TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Kyrill Alexander Baron von Toll 242933IVCM

Towards a Guideline for NIS2 Compliant Security Testing on Digital Twins of Cyber-physical Systems

Master's Thesis

Supervisor: Aleksi Oskar Johannes Kajander
PhD

Tallinn 2026

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Kyrill Alexander Baron von Toll 242933IVCM

NIS2-ga ühilduvate küberfüüsikaliste süsteemide digitaalsete kaksikute turvatestimise suunised

Magistritöö

Juhendaja: Aleksi Oskar Johannes Kajander
PhD

Tallinn 2026

Author's declaration of originality

I hereby certify that I am the sole author of this thesis and that this thesis has not been presented for examination or submitted for defense anywhere else. All used materials, references to the literature, and work of others have been cited.

Author: Kyrill Alexander Baron von Toll

28.05.2026

On the Use of AI and other tools

Throughout this thesis, two AI tools have been used. (i) Grammarly [1] was used for spellchecking and formulation assistance. (ii) ChatGPT [2] was used in four ways. First: as an advanced search engine, to provide an initial understanding of unknown concepts and information, such as protocols and tools previously unknown to the author. Second: for intermittent feedback on the structure and wording of chapters. Third: to generate BibTeX entries for sources where such entries were not provided. Fourth: to translate the title and abstract into Estonian. All suggestions by either tool were manually validated and corrected where necessary.

Author: Kyrill Alexander Baron von Toll
28.05.2026

Abstract

NIS2 is a European Union directive that came into force in December 2022 and, as of the time of writing, remains only partially implemented in several member states. It aims to enhance the cybersecurity and resilience of important and essential entities across the internal market by specifying the requirements they must meet. However, how to fulfill these requirements is not specified, and common industry practices are not always applicable. One such case is the requirement for the assessment and testing of implemented security controls, which can be challenging to conduct on cyber-physical or operational technology systems. As an alternative to testing on the live Cyber-physical or Operational technology systems, it is common to test substitutes such as testbeds or digital twins. These substitutes provide greater flexibility and can save resources, but it remains unclear whether conducting security assessments and tests on them is sufficient to achieve NIS2 compliance. For this reason, this work aims to lay the groundwork for a guideline for NIS2-compliant security testing of digital twins of cyber-physical systems.

The thesis is in English and contains 45 pages of text, 8 chapters, 2 figures, 6 tables.

Annotatsioon

NIS2-ga ühilduvate küberfüüsikaliste süsteemide digitaalsete kaksikute turvatestimise suunised

NIS2 on Euroopa Liidu direktiiv, mis jõustus 2022. aasta detsembris ning mille rakendamine on käesoleva kirjutamise hetkeks mitmes liikmesriigis endiselt vaid osaliselt lõpule viidud. Direktiivi eesmärk on suurendada siseturu oluliste ja elutähtsate üksuste küberturvalisust ja vastupanuvõimet, määratledes nõuded, millele need peavad vastama. Samas ei ole täpsustatud, kuidas neid nõudeid täita, ning tavapärased tööstuspraktikad ei ole alati rakendatavad. Üheks selliseks näiteks on rakendatud turvameetmete hindamise ja testimise nõue, mida võib olla keeruline läbi viia küberfüüsikalistes või operatsioonitehnoloogia süsteemides. Alternatiivina reaalsete küberfüüsikaliste või operatsioonitehnoloogia süsteemide testimisele kasutatakse sageli asendusi, nagu testkeskkonnad (testbeds) või digitaalsed kaksikud (Digital Twins). Need asendused pakuvad suuremat paindlikkust ja võivad säästa ressursse, kuid endiselt ei ole selge, kas nende peal läbi viidud turvahindamised ja testid on piisavad NIS2 nõuetele vastavuse saavutamiseks. Sellest tulenevalt on käesoleva töö eesmärk pakkuda juhiseid küberfüüsikaliste süsteemide digitaalsete kaksikute NIS2-nõuetele vastavaks turvatestimiseks.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 45 leheküljel, 8 peatükki, 2 joonist, 6 tabelit.

Acronyms

AI Artificial Intelligence. 24, 27, 28, 35, 56

BSI Bundesamt für Sicherheit in der Informationstechnik. 20, 41, 46, 57

CAN Controller Area Network. 33, 36, 73

CAV Connected and Autonomous Vehicles. 32

CPS Cyber Physical System. 11, 15–17, 20, 22–24, 27, 28, 30, 31, 34–37, 40, 45, 47, 48, 51–53, 55, 56, 58, 59

CPU Central Processing Unit. 19

CVE Common Vulnerabilities and Exposures. 29

DDoS Distributed DoS. 29, 36

DIN Deutsches Institut für Normung. 27

DORA Digital Operational Resilience Act. 57

DoS Denial of Service. 7, 15, 29, 32, 35

DT Digital Twin. 11, 13, 16–19, 22–27, 29–38, 40, 41, 44–49, 51–56, 58, 59

E-ITS Estonian information security standard. 43

EU European Union. 41, 43, 57, 59

FDIA False Data Injection Attack. 29, 34, 35

HBOM Hardware Bill of Materials. 45

HIL Hardware in the Loop. 25, 29–37, 47

HMI Human Machine Interface. 31, 72–74

IAEA International Atomic Energy Agency. 15

ICS Industrial Control System. 15, 23, 25, 26, 28, 30, 48

IDS Intrusion Detection System. 31, 38

IEC International Electrotechnical Commission. 20, 42, 46, 48, 49, 51, 53, 56

IIoT Industrial Internet of Things. 15, 34, 73

IoT Internet of Things. 24, 28, 34, 73

IP Internet Protocol. 28

ISO International Organization for Standardization. 20, 42, 46, 48, 51–53

IT Information Technology. 15, 18, 20, 32, 35, 41–43, 45–49, 51, 53, 55

MitM Man in the Middle. 29

NASA National Aeronautics and Space Administration. 18

NIS2 securing Network and Information Systems 2. 16, 40–42, 45, 51–53, 55–59

NIST National Institute of Standards and Technology. 8, 20, 26

NIST-CSF NIST-Cyber Security Framework. 20, 24, 37

OPC UA Open Platform Communications Unified Architecture. 30

OpenPGP Open Pretty Good Privacy. 43

OT Operational Technology. 15, 18, 24, 27, 35, 41, 45, 48

PDF Portable Document Format. 40

PLC Programmable Logic Controller. 15, 31, 35, 72–75

PoC Proof of Concept. 25, 27, 36

RCE Remote Code Execution. 33

ROI Return on Investment. 47

SCADA Supervisory Control and Data Acquisition. 26, 72–74

SG Smart Grid. 23

SGCR Smart Grid Cyber Range. 30

STRIDE Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege.. 20, 26, 30

TCP Transmission Control Protocol. 25, 28, 30

TTP Tactics, Techniques and Procedures. 20

VM Virtual Machine. 26, 35, 73, 75

ZTA Zero Trust Architecture. 27

Table of contents

List of figures

List of tables

1 Introduction

The NIS2 directive of the European Union introduced new cybersecurity requirements for essential and important entities and widened the definition of which entities are considered such. One of these requirements is that essential and important entities shall include "policies and procedures to assess the effectiveness of cybersecurity risk-management measures" (Directive (EU) 2022/2555, Art. 21(2)(f) [3]). One way to interpret this article is that some form of active assessment, such as vulnerability scanning, penetration testing, or red teaming, is required. If this interpretation is correct, it leads to a complication. Multiple critical sectors specified in NIS2, such as power, transportation, drinking water, and wastewater, heavily rely on so-called Cyber Physical System (CPS), also known as Industrial Control System (ICS) or Industrial Internet of Things (IIoT). CPSs generally contain both Information Technology (IT) components such as Servers, Network-Routers and PC-Workstations, and Operational Technology (OT) components such as Sensors, Actuators and Programmable Logic Controllers (PLCs). IT components are designed to be versatile and relatively short-lived. They are intended to run many different operating systems, with many more software solutions on top. OT components, on the other hand, are highly specialized and commonly long-lived. They are designed for specific purposes and can be deeply embedded in large, expensive systems, where economic estimation requires decades of runtime for the system to become profitable. Nuclear power plants are an example of such long-lived systems, with 25% of reactors worldwide being older than 40 years and about 68% older than 30 years according to the 2022 International Atomic Energy Agency (IAEA) Safety Reports Series No. 106 [4]. This high specialization and long life cycle of OT components can make them unfit for even the most basic security testing techniques, such as active scanning. An active scan can have unintended Denial of Service (DoS) effects on a CPS [5], which has both security and safety implications. Therefore, the problem is that active security assessments may be required under NIS2 but can also pose risks that operators are unwilling to accept [6]. One solution to this problem is testing on a surrogate system. Historically, such surrogate systems have taken the form of

physical testbeds, in which another system is implemented in hardware for testing. Physical testbeds are the highest-fidelity surrogate, but they have significant drawbacks, particularly their high cost. A more recent approach is to use either fully virtual testbeds or partially virtual, so-called hybrid testbeds. Fully virtual testbeds range from highly sophisticated mathematical models to complete hardware emulation. Hybrid systems integrate hardware components into virtualized environments. Typical approaches for hybrid systems include simulating sensor input or emulating network traffic. A comparison of the benefits and drawbacks of physical-, hybrid-, and virtual-testbeds is shown in Table 1.

Type	**Cost**	**Accessibility**	**Fidelity**	**Flexibility**	**Reproducibility**
Physical	High	Low	High	Low	Low
Hybrid	High	Medium	High	Medium	Medium
Virtual	Low	High	Medium	High	High

Table 1. Tradeoff between physical, virtual, and hybrid testbed design. Generalized from Mumrez et al. [7].

While virtual testbeds offer clear cost benefits, their fidelity remains uncertain. Virtual representations of systems, so-called Digital Twin (DT), have been used with great success across several applications and are considered one of the technologies facilitating Industry 4.0 [8]. Nevertheless, it is unclear whether they are sufficient to meet legal requirements, such as those under NIS2. To address this issue, the following research questions were specified:

RQ: To what extent can DTs be used to conduct NIS2 compliant active security assessments on CPSs?

This leads to the following sub-questions:

- **RQ1: What requirements regarding security assessments are made in NIS2?**
  - RQ1a: Are active security assessments required at all?
  - RQ1b: How should security assessments be conducted according to NIS2?
  - RQ1c: How is the quality of security assessments assured?
- **RQ2: How valid are security assessments conducted on DTs?**
  - RQ2a: What is the risk that a critical vulnerability that would have been found by testing the real CPS is missed when testing on its DT?

  - RQ2b: How different are the results of security tests on DTs compared to tests on the real CPSs?
- **RQ3: What are the benefits of using DTs for security assessments?**
  - RQ3a: Can interrupting the functionality of the CPS for testing be substituted with testing on the systems DT?
  - RQ3b: Is testing on DTs cost efficient?

To answer these questions, an interdisciplinary qualitative survey of technical and legal experts and a semi-systematic literature review were conducted. The results of both were compared against the controls of three relevant standards. The remainder of this thesis is organized as follows: Chapter 2 gives more detailed explanations of terms used throughout the thesis. It also includes brief introductions to relevant laws, official guidelines/their sources, and international and local standards. In Chapter 3, the semi-systematic literature review conducted for this thesis is presented. Chapter 4 describes the research method applied and the execution of the survey. In Chapter 5 the results are presented and discussed in Chapter 6. Finally, Chapter 7 presents concluding remarks and outlines directions for future work.

2 Background

In this chapter, frameworks, organizations, and other terms used throughout the thesis are introduced

2.1 Terms

Throughout the thesis, the terms testbed, DT, and cyber range are used repeatedly and at times interchangeably. To avoid confusion, the definitions and comparisons are given in this section. A simplified visualization of the overlapping use of the terms is provided in Figure 1.

2.1.1 Testbed

For the purposes of this thesis, a testbed is an IT/OT environment built specifically to conduct tests. The test subject can, but does not necessarily have to, be a component of the testbed. E.g., a testbed can be used to assess the effects of cyber-weapons on a power plant without actually attacking a real plant. Testbeds can be physical (built from the same hardware), virtual (have the hardware virtualized), or hybrid, each with its respective benefits and drawbacks (see Table 1).

2.1.2 Digital Twins (DT)

The term DT is very ambiguous, with no unifying definition in the scientific field [8, 9, 10]. The idea was first applied by the National Aeronautics and Space Administration (NASA), which started simulating individual components and systems [10]. The term gained further traction in the context of Industry 4.0, where DTs are proposed for concurrent modeling and optimization of industrial processes[10]. The common baseline for DTs is that they are a virtual/digital representation of a physical system [8]. Some scholars insist that DTs require bidirectional communication with their real counterpart, and the term digital shadow should be used [9]. Others have found that in industry, the term is used more

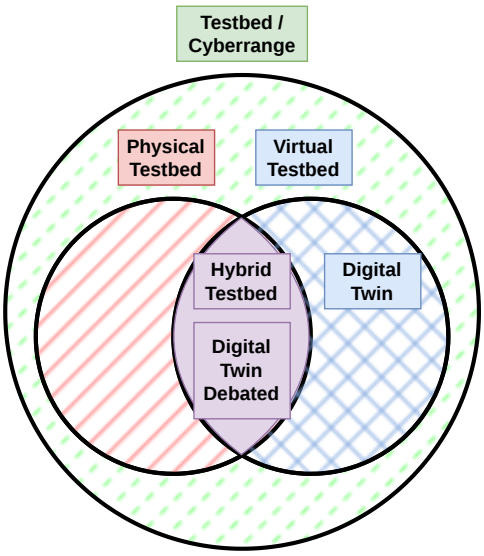

Figure 1. Simplified visualization of the overlapping domains of testbeds, DTs and cyber ranges.

loosely, for example, for virtual models of systems still under design [11]. Another case of ambiguity is the use of integrated hardware. Some scholars insist DTs are purely virtual [9], while others integrate hardware to raise the fidelity of their twin [12]. In the context of cybersecurity, another distinction is how closely the DT simulates the physical state of the components it represents. For example, if a Central Processing Unit (CPU) of a specific architecture is simulated, whether or not to include temperature and cascading effects of temperature [8]. For the purposes of this thesis, particularly the literature review, a broad understanding of the term DT was used. A DT is a virtual representation of a real system.

2.1.3 Cyber Range

Cyber ranges are virtual training environments for cyberattack and defense. They comprise simulated environments, such as single hosts or entire networks, and enable researchers, students, and trainees to test their skills and gain experience in realistic scenarios. Both aggressive and defensive behaviors can be trained in cyberranges, i.e., the trainee might either use a cyber weapon or defend against simulated attacks.

2.2 Organizations and Institutions

In this section, several institutions relevant to and included in this work are named and briefly introduced.

The **International Organization for Standardization (ISO)** is an international, independent, non-governmental organization aiming for internationally aligned standards. They are the source of the ISO 27001 body of standards used in this thesis.

The **International Electrotechnical Commission (IEC)** is also an international organization for standardization with a focus on electronics and electrotechnics. They frequently work in conjunction with the ISO. They are the source of the IEC 62443 body of standards used in this thesis.

The **Bundesamt für Sicherheit in der Informationstechnik (BSI)** is the German Government Office for Information Technology security affairs. They are the source of the IT-Grundschutz body of standards used in this thesis.

The **National Institute of Standards and Technology (NIST)** is the United States Government Office for Standards and Technology. While none of their work is directly used in this thesis, their publications are globally influential, and multiple of them have been cited throughout the literature, such as NIST-Cyber Security Framework or NIST 800-82r3.

There are also numerous frameworks, best practices, and information sources from private entities and corporations. The Purdue model is a framework for safe segregation and architecture of CPS networks. The Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. (STRIDE) framework aids in risk modeling and security assessments by categorizing common attack vectors. The MITRE ATT&CK framework is a large collection of Tactics, Techniques and Proceduress (TTPs) used by threat actors. It helps identify malicious behavior and enables accurate threat simulation.

3 Literature Review

The following literature review strictly focuses on academic works and research. Other sources, such as European Union-published legal documents or relevant industry standards, are discussed in the Background chapter 2.

3.1 Literature review method

The method used in this chapter is a semi-systematic literature review, informed by the PRISMA 2020 statement [13]. The software tool Publish or Perish [14] was used to search Scopus, Semantic Scholar, and Google Scholar. An additional search was made through IEEE Xplore. For Google Scholar, the search was limited to 1,000 results. No other limitations were made. The following search string was used in all searches:

("digital twin*" **OR** "testbed*" **OR** "virtual commissioning"
OR "simulation environment" **OR** "emulation" **OR** "cyber range")
AND
("security testing" **OR** "penetration test*" **OR** "vulnerability scanning"
OR "security assessment"**OR** "fuzz testing" **OR** "red team*" **OR** "threat modeling")
AND
("cyber physical system*" **OR** CPS **OR** "industrial control system*" **OR** ICS
OR SCADA **OR** IACS **OR** "critical infrastructure" **OR** "operational technology"
OR OT **OR** "smart grid" **OR** IIoT **OR** "power plant*")

The number of found Sources per Database/Search Engine is shown in Table 2. The following inclusion/exclusion criteria were applied to the results.

Exclusion Criteria

1. Source not available through the Technical University of Tallinn or the University of Tartu. Only sources that are legally accessible and incur no additional costs were

Database/Search Engine	Number of results
Scopus	127
Semantic Scholar	144
Google Scholar	1000
IEEE Xplore	165
Total Unique results	1146

Table 2. Search results per Database/Search Engine.

considered.

2. Source not available in English or German. Only sources comprehensible to the author without additional translation were considered.
3. Sources Published before 2020. This decision was made to exclude works that rely on outdated technical solutions.
4. Source not focused on security testing or security assessment. Sources focused on topics adjacent to security testing, examples include cybersecurity education and intrusion detection, were not included in this literature review.
5. Source not focused on a CPS.

Inclusion Criteria

1. Sources Published in a peer-reviewed, scientific journal.
2. Conference Papers
3. PhD Theses
4. Technical reports from relevant government organizations

Selection based on the inclusion and exclusion criteria was conducted by the author. Title, abstract, and metadata, such as language, availability, publisher, and date, were manually checked. In total, 82 sources were selected for this review. Because the results were heterogeneous across goals, methods, and performance metrics, they were grouped into 10 categories. These categories, shown in Table 3, represent either the respective field of the CPS (4-9), or, when no specific type of CPS was studied, the study's aim (1-3 & 10). The decision to categorize in this way was made because an early observation was that testbeds and DTs differ significantly by industry.

no.	Category	no.	Category
1	Surveys and literature Reviews	6	Transportation
2	Design, Architecture, and Frameworks	7	Water, Liquids, and Chemicals
3	Testing	8	Agriculture
4	Electrical Power	9	Manufacturing
5	Communication Networks	10	Discussions and Contradictory Sources

Table 3. The ten categories that selected sources were assigned to.

3.2 Surveys and literature Reviews

Six of the selected sources reviewed preexisting work with a broad scope.

Smadi et al. conducted a survey on CPS Smart Grid (SG) testbeds [15]. The authors argue that cyber-physical SGs are particularly vulnerable and require security validation in safe environments. They further set requirements that environments must fulfill and, based on these, identified 19 different SG testbeds, six of which focus on cybersecurity assessment and validation.

Anagnostopoulou et al. bridge the gap between testbeds and DTs by focusing on software tools that emulate ICSs [16]. Their survey identifies 9 such tools and evaluates their respective strengths and weaknesses (for a list of the tools see Table 6). Shi et.al. conduct a similar survey on CPS security in the virtual environment [17]. They differentiate between simulation, emulation, virtualization, and DTs as distinct software types. They identified 75 academic sources, 13 of which focus on DTs, and 32 software tools. They conclude that the digital representation of systems, and DTs in particular, are valid approaches for supporting cybersecurity assessments and research.

In their literature review on DTs, Eckhart & Ekelhart summarize years of their own and other pioneering work on DTs [18]. They provide a comprehensive overview of the varying definitions and applications of DTs and are among the earliest sources to note the usefulness of DT for security and legal compliance. Eckhart was also involved in the Dagstuhl Seminar on CPS security organized by Mora et al. [19]. This Dagstuhl Seminar

was held to bring together experts from around the globe and included several talks on the use of DT for CPS security. Empl et al. conducted a large-scale qualitative systematic literature review on DTs in security operations. Sources identified in their literature review were subsequently categorized by theme, data was extracted, synthesized, and mapped to the NIST-CSF 1.1 [20]. As a result of this mapping, they found that DTs adequately fulfill 4 out of 6 Identify subcategories, 5 out of 6 Protect subcategories, 3 out of 3 Detect subcategories, 4 out of 5 Respond subcategories, and 2 out of 3 Recover subcategories. They further identify strengths and weaknesses of DTs. As strengths, they name asset centrality, (high) fidelity, virtual decoupling, physical/virtual intertwining, and concurrent operation with the real systems' complete lifecycle. The weaknesses they identified are: (i) A lack of standards. (ii) A lack of automated vulnerability response. (iii) A lack of high-level guidance regarding design and access management requirements. (iv) High knowledge requirements. (v) Difficulty in gaining OT/Internet of Things (IoT) network asset visibility. (vi) Difficulty with resilient real-time communication, especially while the DT is under stress. (vii) Cybersecurity and Privacy Concerns. (viii) The need for careful data management and high data quality. (ix) Difficulty maintaining high fidelity and accuracy. (x) Interoperability, integration, and scalability. (xi) (Un-)Explainability of Artificial Intelligence (AI). Finally, Erceylan et al. focus their survey on threat modeling based on DTs [21]. Similar to Empl et al., they mapped the results to the NIST-CSF 1.1 [20] with consideration to the DT modes of operation specified by Dietz et al. [22]. They come to slightly different conclusions, stating that DTs adequately fulfill 4 out of 6 Identify subcategories, 4 out of 6 Protect subcategories, 3 out of 3 Detect subcategories, 3 out of 5 Respond subcategories, and 1 out of 3 Recover subcategories. They conclude that DTs are particularly useful for security testing because they allow testing without interrupting the running CPS and because of the versatility provided by their modes of operation.

These results show that the current research shares concerns about the safety of security assessments on CPS and views DTs as a potential solution. Research into the applicability of this technology for compliance is ongoing, with varying results across the same NIST-CSF framework. This highlights the complications arising from inconsistent definitions of DTs and the lack of formal standardization of the technology.

3.3 Design, Architecture, and Frameworks

While most of the selected sources allude to the designs, architecture, or framework of their work, some have it as their main focus. One recurring complication is the lack of formal standardization or any unified basis of DT design. Sources in this section focus on general testbed or testing-DT design, generating building instructions, and at times Proof of Concept (PoC) implementations. This Section is divided into testbeds, typically focusing on Hardware in the Loop (HIL) implementations, and DTs. Sources are assigned based on their authors' nomenclature

3.3.1 Testbed Design

Gillen et al. lay down fundamental design decisions made for a physical-digital hybrid testbed to assess network intrusion detection solutions in ICSs [23]. They named 5 main objectives for such a testbed: (i) Scenario Realism: the testbed should mimic the real system as faithfully as possible. Replicating with hardware or emulation is preferable where feasible. (ii) Network Realism: Network behavior should be consistent with the real system. The number of connected devices, packet-count, -type, -frequency, etc., should mirror the real system's behavior. (iii) Vulnerability realism: Vulnerabilities of the real system should be represented in the testbed. (iv) Achievable failure: The testbed needs to fail if the real system would. (v) Safety: The testbed must be safe. A PoC was implemented based on a supercomputer cooling system. Their PoC testbed achieved a high fidelity, with less than 1% deviation from the physical counterpart in all metrics but TCP Frames and TCP Bytes, which they attribute to the emulator keeping TCP connections open longer. Validation that the testbed reacts adequately under attacks was not conducted. Green et al. take their design decisions to a higher level of abstraction, focusing on modularity to increase flexibility regarding size and type of ICS [24]. They name 14 aspects to consider in testbed design. These aspects are: Fidelity, Modularity, Diversity, Interoperability, Monitoring and Logging, Openness, Scalability/Extensibility, Flexibility/Adaptability, Repeatability/Reproducibility, Measurability & Measurement Accuracy, Cost-effectiveness, Isolation/Safe Execution, Usability, and Complexity. They further highlight the importance of aligning the testbed with the widely used Purdue model. The authors adjusted their Lancaster University ICS testbed to align with their framework, but they made no mention of validation. AL-Hawareh and Sitnikova use a subset of the Green et al. aspects, namely:

Usability, Fidelity, Heterogeneity, Flexibility and Scalability, Federation, Safety, Reliability, and Resilience, adding User Interfacing, End-to-End Testbed, and Primary Purpose. Their Brown-IIoTbed consists mainly of virtualized components running open-source software. To validate the testbed, AL-Hawareh and Sitnikova conducted several simulated attacks using the STRIDE framework, recognizing that the testbed accurately reacts. They make no evaluation of fidelity. Schmidt et al. focus on the NIST special publication 800-82r3 [25] and build their TROY-testbed in compliance with it [26]. They name Fidelity, Isolation, Repeatability, Measurability, Scalability, and Extensibility as the main requirements for a testbed. To meet these requirements, they replicate a broad ICS network using Virtual Machines (VMs) and other tools (see Table 6). They claim to meet their six stated requirements, mostly by virtue of using VMs, but provide no metrics. Their testbed does match the NIST 800-82r3 guideline. Finally, D'Ambrosio et al. introduce SCASS, a hybrid testbed architecture for Supervisory Control and Data Acquisition (SCADA) systems. They use physical components and virtualize the remaining components using Docker. To validate their architecture, they recreate the EPIC testbed [27] and execute several attacks on the testbed. They further compare their architecture with similar solutions, concluding that the SCASS architecture offers multiple benefits in repeatability, safety, cost savings, fidelity, and scalability. No evaluation of fidelity was made.

3.3.2 Digital Twin Design

Dietz and Pernul were among the first to advocate for DTs for ICS security[22]. They differentiate between several modes of operation that a DT can apply. Additionally to the original Data Analytics and Optimization, it can also be used in simulation or replication mode. According to the authors, simulation mode would be very useful for testing, as tests are easily repeatable and test times can be shortened through accelerated computation. Replication mode, on the other hand, would be useful for observational tasks, such as intrusion detection. In this mode, the DT would receive the same inputs as the real system, and any deviation in the two systems' behaviors would cause alerts. Eckhart et al., following the Dagstuhl seminar mentioned in section 3.3.2, introduced security-enhanced DTs, a framework for comparing DTs with a security focus. For this purpose, they generated a taxonomy of 6 categories with multiple subcategories. (i) Fidelity: focuses on how accurately the DT represents its real counterpart. (ii) Mode of Operation: differs from

that of Dietz and Pernul. It differentiates how the DT is connected to the real system. (iii) Adaptability: how changes to the DT can be implemented. (iv) Repeatability: how accurately the same DT or two DTs of the same system give the same output for the same input. (v) Visibility: how the DT provides information to observers. (vi) Construction: according to which methodology the DT was constructed. Somma et al., like those before, identify criticality of systems and downtime costs as problems for CPS security assessment [28]. They also propose a framework and built a PoC, called MiniCPS, which adequately reacts to tests. They did not evaluate the fidelity of this PoC twin. López-Brea et al. focus on automating the deployment of and security assessment on DTs [29]. They laid out an architectural solution based on two network DTs s, where one DT is used for attack prediction/simulation and the other for impact analysis. They use AI for attack prediction and software-defined networking for automatic deployment. Idika et al. integrate the concept of Zero Trust Architecture (ZTA) into their DT security assessment framework [30]. They highlight that the DTs strength of real-time modeling complements the ZTA paradigm of continuous authentication.

3.4 Testing

Five of the selected sources approached the problem of testing CPS by first answering how to test.

To support industry subject to the Seveso III directive [31], Coppolino et al. introduce the InfraStress-EU framework [6]. They consulted five European operators handling dangerous chemicals and found that the TIBER-EU testing framework published by the European Central Bank [32] is not immediately applicable under their circumstances. Their main contribution, with respect to this thesis, is the implementation of a specific step in system analysis that evaluates the testability of live components. Components that are deemed too critical are subsequently simulated during testing. The effectiveness of their framework was evaluated in a case study in an Italian chemical plant, and the framework has been added to Deutsches Institut für Normung (DIN)-SPEC 91461 [33]. Erkek and Irmak focus on testing the S7 [34] communication protocol[35]. They used several tools (see Table 6) to create a virtualized testbed. They validated the usefulness by conducting several data-alteration attacks, gaining control of the simulated OT component, i.e., turning a lamp on and off.

They conclude that low-cost virtualized testbeds are beneficial for vulnerability assessment of critical systems, as they require no large investments and can provide valuable insights. In a similar fashion, Lazaridis et al. focus on testing the Modbus/TCP protocol [36]. They also use several virtualization tools (see Table 6) to represent a product manufacturing line. They then conduct penetration testing on this testbed, finding all simulated components and intercepting traffic. They conclude by outlining plans to integrate additional communication protocols and then build an automated AI-testing solution for their virtualized testbed.

Kim et al. introduce FirmAE, an emulation solution for the analysis of firmware of embedded IoT devices [37]. They found that prior firmware emulation solutions were insufficient, failing to adequately emulate more than 80% of the Internet Protocol (IP) camera devices they used as their test metric. FirmAE applies what the authors call arbitrated emulation, which focuses on replicating high-level behavior rather than attempting to exactly match the target hardware device. They then conducted a dynamic analysis using a fuzzing approach, identifying several previously unknown vulnerabilities, thereby demonstrating the effectiveness of their approach. Fortino et al. identified another problem with the dynamic assessment of firmware [38]. They claim that emulating firmware as is introduces time constraints because several emulated hardware components, such as timers or bus controllers, slow down dynamic analysis. They propose a solution they call binary rewriting, in which the parts of the firmware that communicate with such components are identified and modified. They provide an example using an ICS humidity sensor but do not provide any metrics. These two sources demonstrate how virtualizing and testing individual components of a CPS can improve security assessment procedures.

3.5 Electrical Power

Sources relating to electricity supply are the most numerous, with 28 selected, of which 22 focus on power grids. This concentration can be attributed to the fact that power grids have long been lucrative targets, especially for state-sponsored actors, as evidenced by attacks such as BlackEnergy 3 [39] and Industroyer 2 [40]. Power grids are heavily reliant on CPSs, and their distributed nature makes them heavily reliant on network- and wireless-communication, creating large attack surfaces. Power-related CPSs also include some of the most safety-critical operations, increasing the risk of testing on live systems.

Examples of such safety-critical operations include managing high-voltage substations and controlling nuclear power plants. Lastly, the concept of smart-grids has long been associated with DTs for performance optimization, indicating preexisting and overlapping expertise between the two fields.

3.5.1 Power grids

Most selected sources focusing on power grids implemented some kind of framework for testbed generation. The identified testbeds are either virtualized [41, 42, 43, 44, 12, 45, 46, 47, 48, 49, 50, 51, 52, 53] or hybrid [54, 55, 56, 57, 58, 59] using a plethora of different tools (see Table 6). Almost all of them validated their testbeds by testing cyberattacks, commonly using DoS/DDoS, Man in the Middle (MitM), and/or False Data Injection Attack (FDIA) attacks. Shitole et al. [43] did not validate their testbed, but laid out plans to do so by testing cyber attacks. Validation through cyber attacks is a reasonable, albeit incomplete, approach. All selected sources have the stated goal to enable cybersecurity research, which includes executing attacks on them, so proving that this goal is met is a basic first step. Furthermore, it shows that the testbed not only reacts appropriately to intended use but also replicates the real system's behavior under unintended circumstances. There are weaknesses, however: (i) some testbeds cannot simulate publicly known vulnerabilities such as Common Vulnerabilities and Exposuress (CVEs), meaning that the attacks must also be simulated, nullifying the comparison. This weakness is stated by Roomi et al. [49] for their own testbed and is suspected in multiple other sources. (ii) Even if some CVE proof-of-concept attack works, this does provide a quantified metric of the testbed's fidelity. Only two of the selected sources conducted their attacks on multiple subjects for comparison. Katuri et al. [58] test on a fully virtual and a HIL hybrid testbed [58]. They present visual representations of their results that do not match. Sen et al. also repeat their attacks in a laboratory and provide only visual representations of their results, but in their case, the results are very similar [52]. Atalay and Angin propose the development of DTs for smart grid security testing and standardization [60]. They compare a series of standards and extract requirements for the DT design. Their result is a hybrid DT with physical, virtual, and decision-making tiers. They give no active recommendations on what tools are adequate for each tier. They conclude that security evaluation can be accomplished by combining this DT with a vulnerability knowledge base, an attack simulation toolset, and a

data analysis module. Bartusiak et al. formulate a methodology for extended gap analysis (also known as compliance checking) in critical infrastructure [61], [62]. They point out that active evaluation techniques, such as penetration testing, pose unacceptable risks for the system. Therefore, a substitute is required to conduct the testing. They validate their methodology through a case study on an actual digital substation in Germany, naming the tools used to represent the system and for penetration testing. Due to the confidential nature of the test results, no metrics are provided. Finally, Mumrez et al. provide valuable insights by conducting a comparative study using the MITRE ATT&CK ICS matrix across three testbeds [7]. They choose the virtual Smart Grid Cyber Range (SGCR) [63], the hybrid KASTEL [64], and the physical EPIC [27] testbeds. They identify 54 exploitable vulnerabilities over all three testbeds combined, 45 for the physical EPIC testbed, 29 for the hybrid KASTEL testbed, and 34 for the virtual SGCR testbed. 16 Vulnerabilities are uniquely exploitable on a single testbed, 9 of which are on Epic, 2 on KASTEL, and 5 on SGCR. They further reaffirm the tradeoffs between testbed designs shown in Table 1.

3.5.2 Power plants

Erkek and Irmak generated a DT of a hydroelectric power plant to conduct cybersecurity testing on [65]. They implemented a Modbus TCP server and slave in Python to function as the DT. Data-alteration attacks were then executed on the twin, enabling continuous writes to CPS components, thereby revealing a vulnerability in the power plant. They did not provide metrics on the twins' fidelity or on whether the attacks were validated. Mathew and Kazi built a HIL testbed of a coal power plant for cybersecurity testing [66]. They analyze the necessary components for modeling and the configuration required. They do not share how they are represented/simulated for all components. STRIDE- based threat assessment is conducted on the model, and attacks are executed against the found vulnerabilities, resulting in system failure on the power plant testbed. Guo et al. propose the implementation of DTs for security assessment with regard to the China GB/T 22239-2019 standard [67]. They point out that standard security measures, such as firewalls and network segmentation, remain valid and widely used, but implementing future technologies will require novel methods to assess their security. Silva et al. introduce the Asherah Nuclear Power Plant Simulator for cybersecurity testing [68]. Their testbed is built in MATLAB/Simulink and can integrate real or emulated hardware components via Modbus or Open Platform

Communications Unified Architecture (OPC UA) communication. They conclude that their work is ongoing, but the testbed is potentially highly valuable for security testing and assessment. Hahn et al. at the Sandia National Labs integrate the Asherah Nuclear Power Plant Simulator in their extended nuclear power plant simulation platform [69]. They extend the work of Silva et al. by incorporating the larger network and providing scalability for the inclusion of PLC via several tools (see Table 6). They conclude by showing that the testbed is operational, but no testing had been conducted at the time of publication. Finally, de Brito and de Sousa introduce an open source testbed, also aimed at nuclear power plants [70]. They also expand on the Asherah Nuclear Power Plant Simulator using multiple open-source tools (see Table 6). To validate their testbed, they conduct a known cyberattack by manipulating the PLC configuration while simultaneously injecting false data into the operators' Human Machine Interface (HMI) dashboard.

3.6 Communication Networks

Like power grids, communication networks are distributed CPS with a large attack surface. All three selected sources in this category have focused on fully virtual DTs. Likely because the number of physical edge devices is significantly lower, and the physical aspect of communications can be accurately simulated, minimizing the need for HIL testing.

Both Rebecchi et al. [71] and Grasselli et al. [72] built DTs of 5G networks based on the ETSI Open Source MANO tech stack. Rebecchi et al. [71] built their DT to be used as a cyber range called SPIDER, and integrated two preexisting testbeds, MATILDA [73], and MOUSEWORLD [74]. They describe the architecture of their cyber range, including their Cyber Risk Assessment Engine, which automatically evaluates risks during exercises. They conclude by noting that future work includes validating SPIDER. In contrast, Grasselli et al. implemented a prototype and validated it by running enumeration and disruption attacks [72]. They further successfully implemented a Snort Intrusion Detection System (IDS) to detect these attacks.

Dauphinais et al. have a slightly different focus, implementing a DT for automated fuzz testing of 5G networks [75]. They validate by including predefined vulnerabilities in their simulated DT, which are identified by the fuzz testing.

3.7 Transportation

Transportation is another frequent category where strong arguments for testing on substitutes can be made. It is the category involving the largest number of untrained human participants, making the security of the systems paramount. It also involves highly dangerous circumstances that cannot be tested safely, such as high-speed autonomous vehicle travel on public roads. Transportation systems are also widely distributed, at times relying on public IT infrastructure, thereby creating large attack surfaces. The rest of this Section is separated by the respective means of transportation.

3.7.1 Railway

Soderi et al. conducted a survey on the cybersecurity of railway systems [76]. They identify the problem of security testing on live systems and the cost of recreating the system with HIL. They propose the same solution as in this thesis: the use of cyber ranges. They specify that the cyber ranges should be able to replicate systems, including their respective circumstances, which they call scenarios. The cyber range solutions they have identified are listed in Table 6. Chen et al. actively compare risk assessment methods used for railway systems [77]. They distinguish three categories of assessment: (i) Qualitative methods, such as the Analytic Hierarchy Process. (ii) Quantitative methods, such as Hidden Markov Models. (iii) Simulation methods, for which they name testbeds and cyber ranges. They conclude that simulation-based methods provide quick and accurate results but are costly in both monetary and human capital terms. One actual implementation of a testbed is the HIL simulation testbed built by Neema et al. for railway systems [78]. Using their prototype, they mirrored the Washington, D.C., metro network. For validation of the testbed, DoS attacks were conducted on Railway-signals, successfully rerouting trains and increasing waiting times. The authors note that their testbed is a fully realistic environment, but provide no metrics to support this claim.

3.7.2 Roads and Road vehicles

Connected and Autonomous Vehicles (CAV), and Cars in general, are a repeating source of interest among the selected sources [79, 80, 81, 82]. Parts of the team of Grasseli et al. expand on their network DT (see Section 3.6) by adapting their model to be useful for CAV

[79]. Both Marksteiner et al. and Wehmer et al. focus on DTs for security testing in the context of the EU implementation of UN Regulation No. 155 [83]. Marksteiner et al. focus on a Python-based virtual implementation to achieve scalable automation [82]. In contrast, Wehmer et al. deliberately focus on emulation and enable HIL integration [81]. This makes their NOTHORG DT labor-intensive to set up, but generates high-fidelity results, which the authors validated through four case studies. Shi et al. built a hardware-based Controller Area Network (CAN)-bus testbed of a 2010 Toyota Camry [80]. They focus heavily on the technical implementation of the testbed and are among the few sources providing a fidelity evaluation. They report that the testbed differs from the replicated vehicle by less than 1% across all tested metrics under normal running conditions. They formulate several attack scenarios, but make no mention of executing any attacks.

The research group around Masi and Sellitto developed a cybersecurity DT architecture methodology and applied it to a road network, which they call the Cooperative Intelligent Transport System, and to a road tunnel system. [84, 85]. They run several attacks on their DT model and repeat the process, incrementally increasing countermeasures, until they reach a secure state at an acceptable risk level. They expressly state that validation through penetration testing of the actual system was not conducted. This would have made the system unavailable, which is unacceptable to the commercial operator of the road network and to the larger population relying on it.

3.7.3 Maritime Vehicles

Lee et al. and Amro and Gkioulos work on HIL testbeds of Autonomous Ships. Amro and Gkioulos propose a versatile testbed with multiple interchangeable tools and components to represent various scenarios [86]. Lee et al. identified the components necessary for autonomous ships and conducted risk modeling using the MITRE ATT&CK framework. They further built a HIL testbed re-hosting/emulating firmware of some components. Penetration tests were conducted, identifying multiple vulnerabilities, including a Remote Code Execution (RCE). They state that their testbed is rudimentary, representing only a subset of a ships network.

Sicard et al. built their testbed for a small surface warship primarily from real-world components, focusing on the ship's energy, propulsion, steering, and weapons systems [87].

They evaluated their testbed using four attack scenarios based on the MITRE ATT&CK for ICS framework [88]. No direct comparison to a real ship was made to validate the results, but the testbed's high fidelity is assumed, as it consists mostly of the same components. In the future, the testbed is expected to be used for further research and for training officers and security specialists.

3.7.4 Airport

Koroniotis et al. introduce SAir-IIoT, a DT replication of an airport with the intent to be offered as a testbed-as-a-service [89]. Their design is hybrid/HIL, as they include 82 IoT devices and 20 virtual machines. The authors compare their testbed with other implementations and find that it is the largest heterogeneous IoT testbed they could identify. To validate the testbed, they collected and provided several telemetry metrics, but no further comparison of these metrics was made. Several attack scenarios were prepared on the testbed, but their execution is not mentioned.

3.8 Water, Liquids and Chemicals

While drinking water was already considered critical under the original NIS [90], the NIS 2 directive included wastewater treatment as a critical concern [3]. While dangerous chemicals also fall under other, stricter regulations, such as the Seveso 3 Directive [31], they are widely used across several critical NIS 2 sectors, including water, health, and agriculture. The handling of such chemicals requires the utmost care, which can also prohibit testing on live systems involving them [6]. This led to five selected sources searching for ways to secure such systems without compromising their safety.

Dietz et al. continue their DT-design work (see Section 3.2) by employing it for security-by-design system testing, including a proof of concept DT of a pressure vessel [91]. Their focus is on the design phase of the CPS; therefore, the DT is a fully virtual simulation of a system that does not yet exist. For security testing, they use the Metasploit framework and custom C++ scripts to execute FDIA and identify vulnerabilities in encryption and authentication. No further information on the implementation of the designed CPS is given. It is unclear whether the found vulnerabilities were mitigated in the actual CPS and whether that mitigation was validated. Also fully virtual, Raj et al. introduce an OpenStack-based

OT testbed to evaluate network-security AI Agents [92]. For this purpose, they conduct two case studies, independently replicating a power grid and a chemical plant. Validation of the testbed was conducted through FDIAs and DoS attacks, which, in the Chemical plant simulation, led to a significant increase in pressure and a subsequent explosion. The authors did not evaluate the fidelity of their testbed. Lin and Liou use a DT to conduct scenario testing on a water pipeline system [93]. They validate via DoS and brute-force attacks, but provide no information on the tools used for the DT, the attacks, or the results. Durazno et al. built VNWTS, a fully virtual drinking-water treatment plant [94]. They validate their approach through attacks, successfully altering the chlorine level to unspecified levels. No fidelity evaluation was made, but they emulate PLCs using vendor-supplied software for this specific purpose. Lastly, Churilla also developed a fully virtual testbed for testing and forensic analysis of engineering workstations [95]. Validation was conducted using DoS, FDIA, and vulnerability-exploitation attacks, ultimately gaining command and control of the workstation. The author concludes that workstation simulation using a VM yields reliable results, whereas the PLC simulation has weaknesses. No fidelity evaluation was done.

3.9 Agriculture

Agriculture is a fundamentally different field from those mentioned before, as at its core it focuses on living beings. Counterintuitively, immediate safety is not the primary problem in security testing for agriculture CPS, as the well-being of livestock is often more of a commercial than an ethical concern. Nevertheless, assessing the security of such systems is necessary, as disruptions in the food supply can have disastrous consequences, ranging from product callbacks to health impacts and famines.

All three identified sources focusing on agriculture testbeds have opted for HIL testbed implementations [96, 97, 98]. It is unclear why no work on fully virtual testbeds was conducted, but there are two running theories: (i) testbeds in this sector are underexplored, (ii) there is less overall IT expertise in farming. Theory (ii) is based on an observation made by Agarwal et al., who note that farmers are seldom IT experts [96]. They further report that this circumstance results in a very flat CPS network topology, commonly lacking segmentation or basic protection mechanisms, such as firewalls. To test the out-of-the-box

security of components, and elicit what security measures are required, Agarwal et al. built a PoC testbed of a dairy farm [96]. Similarly, Brown et al. built a testbed of an Egg production facility [97]. Both sources focus on accurately reproducing the agriculture CPSs and have laid the groundwork for conducting attacks for validation. Neither of them conducted attacks nor evaluated the fidelity of their testbeds. In contrast, Wan et al. conducted extensive validation of their SATB testbed, both manually and with the vulnerability scanner OpenVAS [98]. Their testbed recreates a radio-based network, as it could be used in smart agriculture CPSs.

3.10 Manufacturing

Manufacturing plants were the original inspiration for this thesis. While it can be argued that this sector is less critical than water, agriculture, and electricity, it nonetheless represents a major part of Industry 4.0's economic activity.

Jiang et al. and Francia and Hall focus on virtual DTs of manufacturing systems. Francia and Hall introduce their ongoing work on a bottle-filling system DT, which has a working but unevaluated PoC [99]. Jiang et al. built a twin of a human-robot-assembly system [100]. They validate the twin through attack simulation, identifying a critical vulnerability in password management. Metrics on the validation were not provided, and an evaluation of the twins' fidelity was not made. In contrast to the previous digital approaches, Hariharan et al. implement a HIL testbed of construction machinery for security validation [101]. They validate their approach by conducting DDoS attacks on the CAN-bus, resulting in reaction delays that indicate safety-critical conditions. They do not evaluate the fidelity of their testbed.

3.11 Discussions and Contradictory Sources

As the literature review so far has shown, testbeds and DTs have large potential to enhance the security of CPS. But this potential does not come without a cost, as highlighted by Abdullahi and Lazarova-Molnar, Peck et al., and Gonçalves de Azambuja et al. [102, 103, 104]. Arguments against the use of CPS DTs are:

1. Treasure map: a DT holds as much information on its real twin as possible and in

machine-readable form. This information can be highly dangerous in the hands of malicious actors. Furthermore, DT can at times require privileged access to real systems, meaning it would not only be a treasure map, but also hold the keys.

2. Unproven: as has become evident in this literature review, there has been little work in evaluating the fidelity of security-focused DTs. For now, it remains unclear whether their use will lead to actual security advancements.
3. Lack of standards: there is no apparent consensus on how to build a security-focused testbed or DT. A claim that this literature review also supports, as there are very few sources actively adhering to another source's framework, and among them, no pair adheres to the same one.
4. Increased attack surface: The DT itself can become a point of first entry. Especially, attacks on supply chain simulation software could pose a threat.

3.12 Synthesis of findings

Throughout the literature review, several repeating patterns emerged. Across the sources, there remains little doubt that testing on live CPS is considered a risk, especially for legacy systems. A majority of sources also agree that testing on substitutes can, in principle, improve the cybersecurity posture. There are, however, significant differences of opinion both between sectors and at times within sectors on the preferred type of substitute. In general, it is observed that HIL solutions are preferred for smaller scale CPS and those with the potential of significant physical impact. This can be ascribed to the assumption that the substitute's fidelity increases with higher rates of hardware representation. It can also be observed that the adoption of DTs and other virtual solutions is more prevalent in distributed fields. This, in turn, can be ascribed to the higher scalability of DT solutions. Both fidelity evaluation and, arguably consequently, the applicability for compliance purposes remain underrepresented in the literature. Most sources conduct some kind of baseline fidelity assessment, but there is little actual comparison or even agreed-upon metrics for such comparison. Some initial attempts to assess the compliance applicability of the use of DTs for security assessments have been made. Examples include attempts regarding NIST-CSF [8] and SEVESO [6]. No indicators of acceptance by authoritative organizations were found.

3.13 Common excluded themes

While sources that do not align with the inclusion/exclusion criteria were deliberately excluded, it should be noted that testbeds and DTs can be used to achieve multiple goals simultaneously. Apart from the DTs original purpose of process optimization and life-cycle prediction, the most common uses for DTs not part of the literature review were training, intrusion detection, and honey pots. Educating and training professionals is the primary goal of cyber ranges; it is no wonder that using "cyber range" as a search term yielded many results with this focus. A DT with sufficient fidelity to accurately react to cyber attacks is a prime test subject for learners to practice on. Using DTs as IDS means to monitor differences between digital and real-twin, where any large divergence is an indicator of compromise. Honey pots and honey nets are essentially traps laid out for attackers. They are intended to waste an attacker's time or to collect data on an attacker's behavior to aid counterintelligence.

4 Survey Method

To answer the main Research question (see 1) and Sub-questions (see 1), a two-stage qualitative survey was conducted with experts from technical and legal backgrounds. An exploratory structured questionnaire was used to collect expert input. This input was subsequently analyzed using an inductive qualitative content analysis approach to identify key themes and concerns. These keys and concerns were then applied to authoritative publications identified in the questionnaire with a deductive approach. Finally, the Categories and Themes were compared to identify overlaps and/or gaps. A flowchart of the research process is presented in Figure 2. In the remainder of this chapter, research design, questionnaire design, participant sampling, data analysis procedure, bias and limitations, and ethical considerations are introduced.

4.1 Research Design

Given the legal compliance as the core of this research, a qualitative and exploratory design was deemed the most appropriate approach. The high specificity of the research topic, coupled with its inherent decision-making aspects, led to the choice of expert elicitation as the method. The choice to use questionnaires rather than alternative methods, such as interviews, was made with consideration of the experts' time constraints. Questionnaires provide greater flexibility to both participants and the researchers, as well as greater scalability. Comparing the questionnaire results against authoritative standards was deemed acceptable for validation, as these standards are among the main compliance tools used in practice.

4.1.1 Initial research approach

Initially, the research was designed to follow a Delphi method, which is well-established and useful for predictive research, such as this one. Following the Delphi method was deemed infeasible early on due to limited time. Conducting a limited number of Delphi

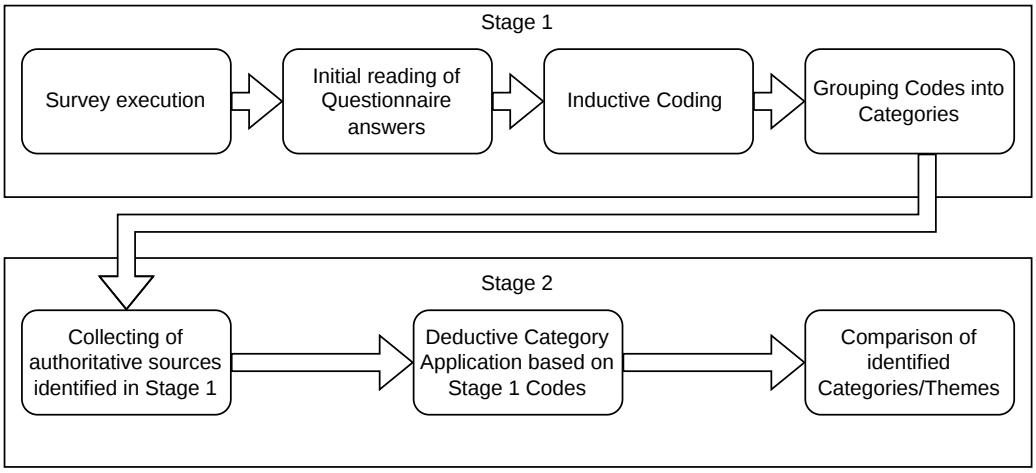

Figure 2. Flowchart of the applied Survey Method.

method iterations, meaning eliciting multiple expert opinions and individually re-presenting them the synthesized findings once or twice, also proved impractical. For this reason, the methodology was adapted as described in Section 4.1, and validation by experts alone was supplemented with active comparison against authoritative standards.

4.2 Questionnaire Design

For the questionnaire design, both content and delivery requirements were considered. For content, the decision was made to limit the number of questions as much as possible to accommodate participants' time constraints. As the survey was exploratory, open-ended questions were used, and participants were encouraged to share as much information as possible. In the end, six questions were asked, focused on: (i) NIS2 compliant security assessments. (ii) Particularities regarding security assessments on CPS. (iii) The possibility to use DTs as substitutes. (iv) The requirements/fidelity of DTs used as substitutes. (v) The cost-benefit estimation for the use of DTs as substitutes. (iv) Any further comments. The questions were ordered to represent the chain of argumentation used in this thesis and identified in several literature sources. Question (i) was chosen as a direct result of the main research question. Questions (ii)-(v) were chosen based on initial literature review results. Question (iv) was added as a best practice and to further encourage the sharing of information. The questions were intentionally simplified and kept broad to accommodate participants' diverse specializations and backgrounds. The topics of each question were briefly introduced to provide context. For the questionnaire, an editable Portable Document

Format (PDF) file was used for distribution across various communication channels. This decision was made deliberately to avoid reliance on a single service provider and to allow participants to take any privacy measures they deem necessary. The final version of the questionnaire is included in the Appendix 8.

4.3 Participant Acquisition and Selection

As NIS2 is a directive, implementation can vary between European Union (EU) member states. To avoid conflicting interpretations arising from differing legal bases, the choice was made to limit participation to experts operating in the Federal Republic of Germany. This decision was also made because the availability of English translations of NIS2 implementations was limited, and the author could read the German implementation in his native language. As the survey was designed to be qualitative, the selection of participants required verifiable expertise. The minimum level of expertise required was 2 years of professional or academic experience in at least one relevant field. The relevant fields specified are Cybersecurity, IT-Security, OT, IT-Law, and EU-Law. Verification of expertise was conducted by cross-referencing information given by participants with publicly available information. The risk of impersonation was acknowledged and deemed acceptable. The acquisition of participants was conducted through purposive sampling in 3 ways: (i) Authors of relevant papers discovered during the literature review. 5 groups with a total of 21 authors were contacted via email. (ii) inquiries towards organizations and experts with public profiles. 9 faculties of German universities with IT-law specialization, 3 Law-firms, and the BSI were contacted via email and through LinkedIn. Another 16 experts were contacted directly via LinkedIn (iii) utilization of private social networks. The questionnaire was shared with former colleagues, classmates, friends, and other social groups. For acquisition method (iii), verification was performed post-facto in the aforementioned manner.

4.4 Data Analysis Procedure

Selected questionnaire responses were analyzed using a qualitative thematic analysis approach. Because research on NIS2-compliant security assessments conducted on DTs remains fragmented, an exploratory inductive coding strategy was adopted for the first stage

of the analysis. This enabled recurring themes to emerge directly from participant responses. The analysis process began with repeated close reading of all responses. During this process, statements and observations were assigned initial codes. Following initial coding, related codes were grouped into categories, which were iteratively compared to reduce overlap and improve internal consistency. Categories were subsequently grouped into themes for the final result discussed in Section 4.1 and Table 4. Furthermore, authoritative standards sources explicitly mentioned in the questionnaire responses were extracted and used in the second stage of analysis.

Three authoritative sources were extracted from questionnaire responses, ISO 27001, IEC 62443, and IT-Grundschutz. For the second stage of the analysis, the derived categories were deductively applied to the controls provided in these three standards. This was done to assess the extent to which existing standards and frameworks met practical concerns. The derived categories were compared against controls provided in the authoritative standards and assigned a value from 0 to 3 based on the degree of agreement between the derived category and the standard control. The values respectively represent:

- 0: Control and category do not fit at all.
- 1: Control introduces measures that implicitly affect the category.
- 2: Control introduces measures that affect the category, but have a different focus.
- 3: Control directly and explicitly affects the category.

The results of the deductive category application are presented in Section 5.2 and Table 5

4.5 Reliability, Bias and Limitations

Several considerations regarding reliability, bias, and limitations have been made throughout the research design process. The inclusion criteria of experts, 2 years of professional or academic experience in one of the related fields, are broad. This decision was made deliberately because the research questions require expertise across several distinct academic and technical fields. While participants with greater experience or specialization across multiple fields might provide more reliable information, recruiting such participants was deemed infeasible. The decision to focus explicitly on the German implementation of NIS2 also imposes limitations on the outcome of this thesis. As mentioned in Section 4.3, this

choice was made for language considerations, as the number of national implementations available in English is limited. The German implementation was chosen for three reasons: (i) German is the author's native language. (ii) The German implementation of article 21.2.f is an almost exact translation of the EU original. (iii) Germany's regulatory and standardization approaches can influence corresponding developments in other EU member states. For example, Estonia partly based its Estonian information security standard (E-ITS) on the German IT-Grundschutz [105]. For these three reasons, focusing on the German implementation was deemed the best option. Including private social networks as pools for participant selection introduces the potential for selection bias. This potential was acknowledged. To reduce potential bias, the expertise of all participants recruited from this group was validated retroactively against the inclusion criteria. Another potential source of bias is that all work in this thesis was conducted solely by the author. Especially the analysis, inductive code extraction, and deductive category application are interpretive work and subject to the author's bias. To reduce inconsistencies, codes, categories, themes, and category-control agreement values were iteratively reviewed multiple times.

4.6 Ethical considerations

Participants were informed of the purpose of the survey and that their responses would be made publicly available. To enable participants to freely share their thoughts, they were given the option to have their information anonymized before inclusion in this thesis. All participants were asked about their willingness to participate in the survey, and gave informed consent. For personal information, only names and self-given descriptions of background and current positions were collected. The survey was distributed via email and the messaging apps Signal and WhatsApp. Outgoing emails were not encrypted because the questionnaire and the survey description were not deemed confidential. An Open Pretty Good Privacy (OpenPGP) key was configured for the author's email account to enable participants to encrypt their answers if they wished. Both messaging apps Signal and WhatsApp provide peer-to-peer encryption [106, 107]. On the author's end, the collected information was stored on four personal devices (a laptop, a desktop computer, an external hard drive, and a mobile phone) and on Tallinn Technical University's email server.

5 Results

In this chapter, the results of both stages are presented. The questionnaire answers are available in the appendix (See 8). Direct and indirect quotes are attributed to the respective participant by ID.

5.1 Inductive Coding and Category Extraction

Coding and Categorization of the questionnaire results revealed four main themes: Assessment practices, CPS-Specific Challenges, Role of DTs, and Economic and organizational factors.

5.1.1 Assessment practices

For assessment practices, participants provided a wide range of specific methods. Five categories were identified. (i) Technical validation measures: Software tools and other processes that provide "[...] measurable verification activities [...] " (P5), such as vulnerability scans (P4, P5), System restoration tests (P1), and security tests (P4, P5, P7). (ii) Organizational evaluation mechanisms: Internal security policy and security governance decisions. This includes documentation of activities (P3), Internal audits (P2, P5), and Reviews (P3, P5, P7). (iii) External validation: Validation and assessments conducted by third parties. External audits (P2, P3, P4) and penetration tests (P4) are explicitly mentioned. (iv) Continuous monitoring and improvement: Assessments need to be ongoing, and findings have to be acted on. This category affects both technical controls, such as "corrective actions" on unsatisfactory backup restoration (P1), and organizational controls, e.g., management reviews (P3, P4, P5, P7). (v) Strategic and Cultural factors: factors that are not necessarily defined in writing, but have a definite impact on the overall security posture. Examples include the risk appetite of decision makers (P5, P6) and the lived security culture among staff members (P3, P5, P6).

5.1.2 CPS-Specific Challenges

Five categories were identified for CPS specific challenges regarding NIS2 compliant security assessments. (i) Operational constraints: CPS specific circumstances, that complicate security assessments, such as up-time requirements or risk of disruptions in operation (P1). (ii) Lifecycle risks: Risk related to the longer lifecycles of CPS hardware, such as disruptions in vendor support leading to the unavailability of certificates and software patches (P5). (iii) Increased impact: Cyber attacks as well as accidents during security assessments have increased impact (P7), as they have the potential to directly cause physical damage. (iv) CPS-related Governance: There still exists a common divide between IT and OT management, which can lead to inconsistent Governance, such as Shadow-IT, i.e., devices that are untracked/unknown to administrators (P6). This extends to interdependencies between the two domains, requiring special attention and documentation, such as an Hardware Bill of Materials (HBOM). (v) Compensating controls: Several participants pointed out that it is possible and common practice to alleviate CPS-specific challenges through compensating controls. This includes controls within the IT domain, e.g., network segmentation and firewalls (P5, P6), as well as assessments in secure, isolated environments (P1).

5.1.3 Role of DT

Participant responses regarding the role of DTs in security assessments were grouped in four categories. (i) Safe Active Validation Capabilities: All seven participants agree that DTs are potentially useful for security assessments. Multiple participant statements support the hypothesis that DTs provide safe and flexible testing environments (P1, P2, P3, P4, P6). Another strength repeatedly highlighted is that building a DT requires accurate knowledge of its real-system counterpart. This makes DTs effective for risk management during the design phase (P2, P7) and, if implemented well, it serves as a detailed index of system components (P5). (ii) DT types: all participants pointed out that the type of DT used has to be adapted to its relative use-case. (iii) Fidelity: similar to the type of DT, participants also pointed out that the fidelity of DTs is use case dependent. Some participants state that lower fidelity can be acceptable depending on the use case and risk appetite (P6, P7). Several participants rejected the simplified representation of fidelity in the survey and provided more detailed descriptions of which system components can and should be modeled with

the highest possible accuracy. Participant 3, with a legal background, estimated that a 70+% fidelity level could be sufficient for compliance purposes. (iv) DT Limitations: Several limitations were mentioned by participants. Participant 4 noted that DTs neglects human vulnerabilities in the system. Participant 5 stated that third-party software and hardware cannot always be modeled and may have to be treated as black boxes.

5.1.4 Economic Considerations

One of the main changes from NIS1 to NIS2 is the explicit inclusion of the cost of non-compliance. Whether DTs will see real-world use for NIS2 compliance depends on the cost-benefit analysis conducted by affected parties. Three categories of economic considerations have been identified in the participants' answers. (i) Cost: the majority of participants doubt that DTs are cost-effective for security assessments (P1, P2, P3, P4, P6). They point not only to hardware costs but also to the cost of the talent and time required to build and maintain DTs. (ii) Compliance vs. Security: Participants mentioned that the focus on NIS2 compliance is misguided. Participant 2 calls NIS2 "[...] a relatively simple and crude approach to enhancing cybersecurity." Participant 3 suggests "[...] DT not being worth the cost "just" for achieving NIS2 compliance [...]" (iii) DT alternatives: Multiple participants pointed to alternative security assessment options as preferable (P2, P7).

5.2 Deductive Category Application results

Deductive category application of the categories identified in phase 1 was conducted on three bodies of standards, ISO 27001:2022, IEC 62443, and the IT-Grundschutz by the German BSI. Specifically, the categories were applied to controls provided in these standards. An overview of the phase 2 results is provided in table 5.

5.2.1 Assessment Practices

All three standards are rather strong in all five categories of the assessment practices theme. ISO 27001 is less specific overall, requiring that certain security levels be met but not specifying how to achieve them. This makes it comparatively weaker on technical validation. IEC 62443 is very strong on technical validation and strong on the other categories as well. The only category where it is comparatively weaker is External Validation. While the

Theme	Categories	Example Code
Assessment practices	Technical validation measures	Penetration testing
	Organizational evaluation mechanisms	Internal Audits
	External validation	External audits
	Continuous monitoring and improvement	Feedback loops
	Strategic and Cultural factors	Risk-appetite
CPS-Specific Challenges	Operational Constraints	Safety constraints
	Lifecycle Risks	Legacy Systems
	Increased Impact	Safety risks
	CPS-related Governance	Shadow IT
	Compensating Controls	Network Segmentation
Role of DTs	Safe Active Validation Capabilities	Risk-free testing
	DT Types	HIL DT
	Fidelity	Use-Case Dependent
	DT Limitations	Limited Human consideration
Economic Considerations	Cost-Benefit Evaluation	ROI considerations
	Compliance vs Security	DT not required for NIS2
	DT alternatives	Real-life exercises

Table 4. Themes and Categories Identified in Stage 1 of the Survey.

standard is comprehensive in its audit requirements, it is the only one that does not explicitly require regular external audits. IT-Grundschutz builds on ISO 27001 by providing more concrete measures to be taken to achieve control requirements.

5.2.2 CPS-Specific Challenges

For the CPS-Specific Challenges IEC 62443 is strong across all categories, while the IT-Grundschutz is partially-, and ISO 27001 is severely lacking. This outcome was expected, as IEC 62443 has an active OT focus, while ISO 27001 is strictly focused on IT and organizational implementations. IT-Grundschutz provides seven OT focused Bausteine (groups of controls), but lacks the specificity on operational constraints and the quantity of compensating controls that IEC 62443 provides.

5.2.3 Role of Digital Twins

All three standards are weak on the theme of DTs and the underlying problems and categories associated with it. IEC 62443 provides exhaustive information on the execution of passive validation, including risk-assessment matrices and risk measurement formulas, but active validation is rarely mentioned. Furthermore, none of the standards explicitly mentions the potential dangers of active validation or testing. IEC 62443 acknowledges that such validation ought to be conducted in maintenance periods and only in agreement with asset- and process-owners. The IT-Grundschutz goes a step further and allows for validation of ICS component software updates in safe testing environments [108][IND.2.1.A1]. This is the closest mention to the use of DTs or testbeds in general for security assessments identified in the three standards. The IT-Grundschutz also provides the Bausteine "SYS.1.5 Virtualisierung" [109] and "SYS.1.6 Containerisierung" [110] for virtualization and containerization of services, which are loosely related to DT-types.

5.2.4 Economic Considerations

Emphasis on the three identified categories under the theme of Economic considerations varies between the three standards. Coverage in the IT-Grundschutz is weakest, with no explicit mention of cost-benefit evaluations. ISO 27001 and IEC 62443 require some extent of cost-benefit analysis, with IEC 62443 dedicating a whole subchapter on business rationale, including recommendations for building a convincing business case for cybersecurity

[111][IEC 62443-2-1 A.2.2]. For the tradeoff between compliance and security, all three standards leave room for interpretation. For example, the IT-Grundschutz formulations intentionally and in the majority of cases use "should", while "must" is reserved for rare instances. For DT alternatives, all standards provide compensating controls as specified under category 2.5, with IEC 62443 adding the aforementioned extensive guidance on passive risk assessment and threat modeling.

Theme	Category	ISO/IEC 27001	IEC 62443	IT-Grundschutz	Key Insight
Assessment Practices	1.1 Technical Validation	●●○	●●●	●●●	IEC more CPS-oriented validation
	1.2 Organizational Evaluation	●●●	●●●	●●●	Strong governance in all three
	1.3 External Validation	●●●	●●○	●●●	ISO more certification-driven
	1.4 Continuous Monitoring	●●●	●●●	●●●	Strong but not CPS-specific in ISO
	1.5 Strategic & Cultural	●●●	●●●	●●●	Strong governance emphasis
CPS-Specific Challenges	2.1 Operational Constraints	●○○	●●●	●●○	Major gap in ISO
	2.2 Lifecycle Risks	●●○	●●●	●●●	IEC explicitly lifecycle-driven
	2.3 Increased Impact	●○○	●●●	●●●	Physical impact mainly in IEC
	2.4 CPS Governance	●●○	●●●	●●●	IEC more detailed structure
	2.5 Compensating Controls	●●○	●●●	●●○	Strong concept only in IEC
Role of Digital Twins	3.1 Safe Active Validation	○○○	○○○	●○○	IT-G mentions safe testing environments
	3.2 DT Types	○○○	○○○	●○○	IT-G discusses Virtualization
	3.3 Fidelity	○○○	○○○	○○○	Not addressed
	3.4 DT Limitations	○○○	○○○	○○○	Not addressed
Economic Considerations	4.1 Cost-Benefit	●●○	●●●	●○○	Implicit in all standards
	4.2 Compliance vs Security	●●○	●●○	●●○	Present but not explicit
	4.3 DT Alternatives	●○○	●●○	●○○	Testing exists but limited for CPS

Table 5. Results summary of category coverage across ISO/IEC 27001, IEC 62443, and IT-Grundschutz.

6 Discussion

The aim of this thesis is to elaborate whether or not DTs are valid and reasonable tools to achieve NIS2 compliance. Specifically to support the assessment of the effectiveness of cybersecurity measures. For this purpose, the remainder of this chapter synthesizes results from phases 1 and 2 of the survey and the literature review. Subsequently, alignment with the three identified standards, ISO 27001, IEC 62443, and IT-Grundschutz is discussed. A critical reflection on the survey's execution, limitations, and validity is conducted. Finally, the main and sub-research questions are reexamined and answered.

6.1 Survey findings and Literature Review results

Through the survey, four main themes of concerns regarding the use of DTs for NIS2 compliance were identified. (i) Assessment practices, (ii) CPS-Specific Challenges, (iii) Role of DTs, and (iv) Economic considerations.

6.1.1 Assessment practices

Throughout the literature review, many approaches to using DTs for the technical validation of cybersecurity measures were identified. The overwhelming focus is on building environments for active testing of CPSs, which otherwise could not be tested for numerous reasons. This is the gap in security assessments on CPS that has been identified leading up to this thesis as well. The multitude of identified sources that build and validate technical solutions leads to the conclusion that DTs and other testbeds can reasonably be used for active security assessments. The survey, however, revealed that technical validation measures may be necessary but are not sufficient for achieving NIS2 compliance. Survey participants repeatedly emphasized requirements for the larger socio-technical system that are not adequately represented in the technical implementations reported throughout the literature review.

6.1.2 CPS-Specific Challenges

There is overwhelming agreement between the literature review sources and the survey results regarding the challenges posed by CPS environments. While some sources predict greater resilience of CPS components in the future, survey and literature review results indicate that active testing of CPS is risky. This risk is driven by operational constraints, lifecycle implications, and the increased impact of mistakes and may be unacceptable, depending on a given organization's risk appetite. Literature review sources and survey participants further agree that DTs/testbeds are an adequate solution to mitigate this risk. Furthermore, there are cases in which the only options are to use such a substitute or to rely entirely on compensating controls. DTs do not provide solutions for Governance issues and life-cycle risks, but they can provide insight into both by serving as an inventory of all CPS components, thereby supporting visibility.

6.1.3 Role of Digital Twins

As noted earlier, survey participants and the literature review agree on the potential of DTs as a safe testing environment. Opinions regarding required fidelity vary, with some sources developing highly theoretical models that represent only specific parts of a system. However, there is a consensus that the DT design is use-case-dependent, and that fidelity can be adjusted accordingly. As an extreme example, it is possible, albeit of questionable utility, to fully emulate every component and thereby achieve a near-perfect fidelity. Nevertheless, significant weaknesses remain. As identified in Subsection 6.1.1, DTs are not a one-size-fits-all solution for NIS2 compliant, as they can only support organizational and strategic measures taken. Furthermore, they require additional cybersecurity vigilance because they introduce new attack surfaces, as discussed in Chapter 3.11. Finally, there is an undeniable lack of standardized DT implementations, as noted by several sources and evidenced by the abundance of different frameworks. While standards on the design and use of DTs exist, such as ISO 23247 [112], adoption is ongoing, and there is little emphasis on the use for active security assessments. This culminates in the conclusion that, while there is agreement on the theoretical potential of DTs, their use for cybersecurity purposes lacks the practical maturity required to be considered a viable compliance tool.

6.1.4 Economic Considerations

Throughout the research period, economic considerations appeared to be the main hindrance to the adoption of DTs as cybersecurity tools. While several sources in the literature review propose low-cost DT solutions, e.g. [98, 113], the majority of sources indicate that DTs are, or are at least perceived to be, significantly resource-consuming. This aligns with the survey results, in which all participants expressed doubt about the cost-effectiveness of DTs as cybersecurity tools, with common concerns being acquisition costs, initial development effort, and ongoing cost for skilled DT-engineers.

6.2 Standards integration

As shown in Table 5, all three examined standards, ISO 27001, IEC 62443, and IT-Grundschutz, sufficiently meet the requirements for assessment practices identified through the survey. It can be assumed that they therefore also sufficiently guide implementing parties towards achieving NIS2 compliance. The fact that ISO 27001 is comparatively weaker on CPS-Specific Challenges can be explained by its inherent IT focus, and does not diminish its usefulness in regard to NIS2 compliance. The fact that all three standards have very weak coverage of the DT theme leads to several conclusions. First and foremost, it supports the claim made in subsection 6.1.3, that there is a lack of standardization regarding DTs and their use in cybersecurity. Conversely, this also leads to the conclusion that complications arising from active security testing on CPS are not perceived as urgent problems by at least three of the involved standards organizations. There is no discernible statement prohibiting the use of DTs in any of the three standards, but also no mention of substitute testing subject apart from the vague mention of safe testing environments in IT-Grundschutz Baustein IND.2.1.A1 [108]. Therefore, it can be concluded that DTs are permissible tools, but in no way are they, or the function they fulfill, required.

6.3 Critical Reflections

There are several weaknesses in the academic work presented in this thesis, both in the conducted survey and in the literature review and its identified results. This subsection acknowledges these weaknesses and discusses the limitations they impose on the results.

6.3.1 Soft definition of relevant terms

As mentioned multiple times throughout the thesis, the definition of the term DT is highly debated [114]. In the context of this thesis, this circumstance had positive and negative impacts. The positive effect is that the term leaves room for interpretation, which matches both the regulation and the three identified standards. This means that many cybersecurity solutions already in place, e.g., MATLAB simulations, could reasonably be called DTs. This, however, is also the greatest flaw in the soft definition. Throughout the literature review, vastly different systems have been referred to as DTs. Systems ranging from purely theoretical models, which essentially only consist of mathematical equations, to complete replicas with almost full emulation of all components [113]. Furthermore, some systems that could reasonably be called DT have been presented as testbeds or cyberranges instead. This led to an early decision to include both terms in the literature review search string, which, in turn, inflated the results and further reduced the search objective's specificity. These circumstances significantly complicated the derivation of meaningful conclusions for the stated research questions.

6.3.2 Survey execution and survey participation

The Survey did not achieve the stated goals for participant numbers and background distributions. The initial aim was to recruit at least 5 participants for each of the technical and legal domains. With 6 participants with technical backgrounds and 1 with a legal background, this requirement was not met, which negatively affects the survey's validity. There seem to be several, at times opposing, reasons for the low participant turnout. First, time constraints for the contacted experts. Most contacted experts did not respond. Out of the actual respondents, most stated that they do not have the capacity to participate. This constraint was anticipated, and the survey was designed to be as short as possible, but the measures taken apparently proved insufficient.

Another anticipated problem was the subject's technical complexity. The survey was intentionally designed to minimize technical complexity. For each question, the relevant terms were introduced in a simplified manner, and participants were strongly encouraged to contact the author with any questions. The design choice to simplify terms had a negative impact in two directions. On the one hand, simplification was apparently

insufficient. Several legal scholars from the IT-Law faculties of four different German-speaking universities responded that they do not feel qualified to answer the questions. On the other hand, the questions were deemed as over-simplified by technical experts. Participant 6 states: "Difficult to be more specific with such questions [...]". Another contacted researcher and author of a source in the literature review responded:

> Simulation and emulation, for example, are fundamentally different concepts and will dictate the "value" and fidelity of the DT. Since it's implementation- and use case-dependent, I am afraid I am not able to fully answer the survey.

(E-mail correspondence, February 2026)

Finally, there appeared to be a degree of guardedness in the provision of "legal advice" by both technical and legal experts. The author tried to alleviate any concerns in this regard by clearly stating that all answers were to be treated strictly as personal opinions. Furthermore, participants were given the option to have their answers anonymized. Nevertheless, several contacted experts responded that they are unable or unwilling to participate, citing that their answers could be construed as official statements. One argument in this regard was that the survey requests background information about the participant. This request was intended only for the validation of expertise, but the intent could have been expressed more clearly.

6.3.3 Answering the Research Questions

The full research questions are stated in Chapter 1, but as a short summary, they are:

- RQ: To what extent can DTs be used to conduct NIS2 compliant active security assessments on CPSs?
- RQ1: What requirements regarding security assessments are made in NIS2?
- RQ2: How valid are active security assessments conducted on DTs?
- RQ3: What are the benefits of using DTs for security testing?

RQ3: Multiple benefits of DTs for security testing have been identified throughout the research. The literature review shows that they fulfill the promise of being an environment to test in and on, with a significant reduction in safety risks. This further includes strengths such as the repeatability of tests under identical conditions and the scalability of the test

environment to the exact needs of the test. For RQ3a, the three standards used for phase 2 do not allow the conclusion that interruption of the real-life CPS can be completely foregone by testing on the DT. Especially IEC 62443 requires tests to be conducted during maintenance phases on the CPS. However, academic works such as Coppolino et al. [6], standards like DIN SPEC 91461 [33], and some survey responses (Participant 7) indicate that DTs are adequate alternatives when CPS interruptions are unsafe or unfeasible. For RQ3b, whether using DTs is cost-efficient, the conclusion reached in this thesis is not yet. All survey participants showed some level of skepticism, and for good reason. While some sources in the literature review reported low-cost implementations of DTs, there was no example of a low-cost, high-fidelity solution suitable for security assessments.

RQ2: Research Question 2 focuses on the fidelity of DTs as the main basis of certainty that the results of a security assessment conducted on a DT are accurate. Among the stated research questions, the answer to this one is the least clear. While the majority of the literature review sources included some form of fidelity validation mechanism, many did not specifically quantify fidelity. Many sources that built actual CPS substitutes opted to conduct cyberattacks against them as validation. Showing that the substitute behaves similarly under both normal conditions and stress can indeed support a claim of high fidelity, but very few sources present concrete data on attack outcomes. Even fewer sources conducted the attacks on both real CPS and substitute to provide comparable data. This lack of interpretable data and clear answers aligns with survey participants' responses to survey question 4. Furthermore, it complicates answering RQ2b: How different are the security test results for DT compared with the real system? Here, different literature review sources yield results ranging from very similar to not similar. This research was unable to answer RQ2a: what is the risk that a critical vulnerability is missed on DT but would have been found in the real system? Primarily because of the aforementioned lack of data directly comparing security assessment results between DT and the real system. Secondly, due to recent and drastic changes in the discovery of critical vulnerabilities supported by AI, indicating a paradigm shift [115, 116].

RQ1: Opinions and information regarding the necessity of and requirements for active testing for NIS2 compliance vary. While three survey participants named active assessment procedures (Participants 4, 5, and 7), others named alternatives that can potentially be

conducted passively. None of the three analyzed standards specifically requires active security assessments. Furthermore, a webinar by the BSI attended by the author [117] emphasized the reporting requirements in NIS2, naming documentation of cybersecurity measures and implementation of reporting processes as priorities for compliance. Finally, a direct comparison with other EU Laws, such as the Digital Operational Resilience Act (DORA) regulation [118], shows that there are cases in which requirements for active testing are made specifically. The combination of these results leads to the conclusion that NIS2 does not strictly require active security assessments, answering RQ1a and rendering RQ1b and RQ1c obsolete.

7 Conclusion

This thesis set out to answer the main research question: To what extent can DTs be used to conduct NIS2 compliant active security assessments on CPSs? As shown by multiple examples in the literature review, and reaffirmed by phase 1 of the survey, DTs can be useful tools for cybersecurity in general and active security assessments in particular. They provide a safe testing environment, which is particularly useful in circumstances where active security testing on running production is unsafe and interrupting production is unfeasible. This supports achieving a better cybersecurity posture and, therefore, NIS2 compliance. However, the research presented in this thesis concludes that active security assessments are not strictly required for NIS2 compliance. Furthermore, it has been shown that the use of DTs for security assessments remains underexplored, especially in commercial use. As DTs are highly context-and use-case-dependent, there is no unified standard for fidelity or cost-benefit evaluation. This, in turn, weakens the business argument for the use of DTs for cybersecurity. In conclusion, despite the limitations discussed in Chapter 6, the findings of this thesis suggest that the use of DTs for active security validation is not necessary to achieve NIS2 compliance. Existing standards and established assessment methods already provide viable compliance pathways. However, DTs may provide substantial benefits in safety-critical CPS environments where active testing on live systems is infeasible or unacceptable. In such contexts, DTs can contribute to a cybersecurity posture exceeding minimum compliance requirements. Operators should evaluate the adoption of DT on a case-by-case basis. Given the technology's current maturity, its implementation should be driven by risk-mitigation requirements and operational constraints, not by compliance alone.

7.1 Future Research and Recommendations

The main obstacle to the use of DTs for compliance purposes, as identified in the literature review, is the lack of comprehensive fidelity evaluations of prototype testbeds and DTs. A comparative analysis of achievable fidelity thresholds across different DT approaches

could be very beneficial in this regard. The lack of a unified definition of DTs and the absence of standardized design methodologies used in practice complicate both research comparability and practical adoption. Future work should therefore focus on developing shared terminology and standardized DT architectures for cybersecurity applications. Furthermore, greater focus on the fidelity evaluation of DTs used in cybersecurity research would be beneficial. This could not only provide clearer validation of research results but also increase trust in the technology and consequently improve practical applicability. Lastly, the conclusions drawn in this thesis are subject to change. The current immaturity of DTs for cybersecurity use, as well as uncertainties on the level of enforcement of NIS2 in Germany and the EU, are ongoing limitations. As DT technologies and NIS2 implementation practices mature, larger-scale studies may provide more representative insights into the practical role of DTs in CPS security assessments.

8 Acknowledgments

The author would like to express sincere gratitude to Dr. Aleksi Kajander for the guidance, feedback, and support provided throughout this thesis.

These thanks extend to Shaymaa Mamdouh Khali, likewise for the feedback provided at intermittent stages of the Thesis.

The author also wishes to thank Tallinn University of Technology and the University of Tartu for providing an inspiring and positively challenging academic environment during the master's studies.

Lastly, the author gratefully acknowledges financial support from the German Academic Exchange Service (DAAD), which enabled completion of the master's studies.

References

[1] Grammarly Inc. *Grammarly*. AI-based writing assistant used for grammar and style checking. 2026. URL: https://www.grammarly.com/.

[2] OpenAI. *ChatGPT (GPT-5.3)*. Large language model used for language assistance, explanations, and code suggestions. 2026. URL: https://chat.openai.com/.

[3] European Parliament and Council of the European Union. *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)*. https://eur-lex.europa.eu/eli/dir/2022/2555/oj. Official Journal of the European Union, L 333, 27 December 2022, pp. 80–152. 2022.

[4] International Atomic Energy Agency. *Ageing Management and Long Term Operation of Nuclear Power Plants: Data Management, Scope Setting, Plant Programmes and Documentation*. Tech. rep. Safety Reports Series No. 106. Vienna: International Atomic Energy Agency, 2022. URL: https://www.iaea.org/publications/14693/ageing-management-and-long-term-operation-of-nuclear-power-plants-data-management-scope-setting-plant-programmes-and-documentation.

[5] Ondrej Pospisil et al. "Active Scanning in the Industrial Control Systems". In: Nov. 2021, pp. 227–232. DOI: 10.1109/ISCSIC54682.2021.00049.

[6] Luigi Coppolino et al. "A framework for Seveso-compliant cyber-physical security testing in sensitive industrial plants". In: *Computers in Industry* 136 (2022), p. 103589.

[7] Aneeqa Mumrez et al. "Comparative study on smart grid security testbeds using mitre att&ck matrix". In: *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE. 2023, pp. 1–7.

[8] Matthias Eckhart and Andreas Ekelhart. "Digital twins for cyber-physical systems security: State of the art and outlook". In: *Security and Quality in Cyber-Physical Systems Engineering: With Forewords by Robert M. Lee and Tom Gilb* (2019), pp. 383–412.

[9] Werner Kritzinger et al. "Digital Twin in manufacturing: A categorical literature review and classification". In: *Ifac-PapersOnline* 51.11 (2018), pp. 1016–1022.

[10] Philip Empl et al. "Digital twins in security operations: State of the art and future perspectives". In: *ACM Computing Surveys* 58.1 (2024), pp. 1–38.

[11] Luiz Fernando CS Durão et al. "Digital twin requirements in the context of industry 4.0". In: *IFIP international conference on product lifecycle management*. Springer. 2018, pp. 204–214.

[12] Nandha Kumar Kandasamy et al. “An electric power digital twin for cyber security testing, research and education”. In: *Computers and Electrical Engineering* 101 (2022), p. 108061.

[13] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *bmj* 372 (2021).

[14] Anne-Wil Harzing. *Publish or Perish*. Computer software. Available at https://harzing.com/resources/publish-or-perish. 2007.

[15] Abdallah A Smadi et al. “A comprehensive survey on cyber-physical smart grid testbed architectures: Requirements and challenges”. In: *Electronics* 10.9 (2021), p. 1043.

[16] Argiro Anagnostopoulou et al. “From Plant to Lab: Industrial Emulation Tools for Real-World Security Testing in Industrial Control Systems.” In: *SECRYPT*. 2024, pp. 13–25.

[17] Lei Shi, Shanti Krishnan, and Sheng Wen. “Study cybersecurity of cyber physical system in the virtual environment: a survey and new direction”. In: *Proceedings of the 2022 Australasian Computer Science Week* (2022), pp. 46–55.

[18] Matthias Eckhart and Andreas Ekelhart. “Digital twins for cyber-physical systems security: State of the art and outlook”. In: *Security and Quality in Cyber-Physical Systems Engineering: With Forewords by Robert M. Lee and Tom Gilb* (2019), pp. 383–412.

[19] Alvaro Cárdenas Mora et al. “Digital twins for cyber-physical systems security (Dagstuhl Seminar 22171)”. In: *Dagstuhl Reports* 12.4 (2022), pp. 54–71.

[20] National Institute of Standards and Technology. *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*. Tech. rep. NIST CSF 1.1. Accessed: 2026-02-25. U.S. Department of Commerce, 2018. URL: https://www.nist.gov/cyberframework.

[21] Gizem Erceylan, Aida Akbarzadeh, and Vasileios Gkioulos. “Leveraging digital twins for advanced threat modeling in cyber-physical systems cybersecurity: G. Erceylan et al.” In: *International Journal of Information Security* 24.3 (2025), p. 151.

[22] Marietheres Dietz and Gunther Pernul. “Unleashing the digital twin’s potential for ics security”. In: *IEEE Security & Privacy* 18.4 (2020), pp. 20–27.

[23] Robert E Gillen et al. “Design and implementation of full-scale industrial control system test bed for assessing cyber-security defenses”. In: *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE. 2020, pp. 341–346.

[24] Benjamin Green et al. “{ICS} testbed tetris: Practical building blocks towards a cyber security resource”. In: *13th USENIX workshop on cyber security experimentation and test (CSET 20)*. 2020.

[25] National Institute of Standards and Technology. *Guide to Operational Technology (OT) Security*. Tech. rep. NIST Special Publication 800-82 Revision 3. Gaithersburg, MD: U.S. Department of Commerce, 2023. DOI: 10.6028/NIST.SP.800-82r3. URL: https://doi.org/10.6028/NIST.SP.800-82r3.

[26] Henry Schmidt, Gideon Sutterfield, and Chris Farnell. “Addressing Cybersecurity Data and Workforce Scarcity with TROY: Testbed for Resilient Operational sYstems”. In: *2024 IEEE Design Methodologies Conference (DMC)*. IEEE. 2024, pp. 1–8.

[27] Sridhar Adepu, Nandha Kumar Kandasamy, and Aditya Mathur. “EPIC: An Electric Power Testbed for Research and Training in Cyber Physical Systems Security”. In: *Computer Security*. Ed. by Sokratis K. Katsikas et al. Cham: Springer International Publishing, 2019, pp. 37–52. ISBN: 978-3-030-12786-2.

[28] Alessandra Somma et al. “A cyber digital twin framework to support cyber-physical systems security”. In: *2023 IEEE Smart World Congress (SWC)*. IEEE. 2023, pp. 1–10.

[29] Juan Tamboleo López-Brea et al. “Dynamic Deployment and Security Assessment of Resilient Services over Digital Twins”. In: *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE. 2024, pp. 836–841.

[30] Chima Nwankwo Idika et al. “Digital Twin-Enabled Vulnerability Assessment with Zero Trust Policy Enforcement in Smart Manufacturing Cyber-Physical System International Journal of Scientific Research in Computer Science”. In: *Engineering and Information Technology* 9.6 (2023).

[31] European Parliament and Council of the European Union. *Directive 2012/18/EU on the control of major-accident hazards involving dangerous substances (Seveso III)*. Official Journal of the European Union, L 197, 24 July 2012. Accessed: 2026-02-28. 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012L0018.

[32] European Central Bank. *TIBER-EU Framework: How to implement the European framework for Threat Intelligence-based Ethical Red Teaming*. Tech. rep. Accessed: 2026-02-28. Frankfurt am Main: European Central Bank, 2018. URL: https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework.en.pdf.

[33] DIN Deutsches Institut für Normung e.V. *DIN SPEC 91461:2021-12 – Stresstest der Resilienz von kritischen Infrastrukturen, die cyber-physischen Bedrohungen ausgesetzt sind*. DIN SPEC DIN SPEC 91461:2021-12. ICS 13.200. Berlin: DIN Deutsches Institut für Normung e.V., Dec. 2021.

[34] *ISO Transport Service on top of the TCP Version: 3*. RFC 1006. May 1987. DOI: 10.17487/RFC1006. URL: https://www.rfc-editor.org/info/rfc1006.

[35] Ísmail Erkek and Erdal Irmak. “Evaluation of SCADA Test Beds and Design of a New Software-Based Test Bed”. In: *2023 11th International Conference on Smart Grid (icSmartGrid)*. IEEE. 2023, pp. 1–8.

[36] George Lazaridis et al. “Securing modbus tcp communications in i4. 0: A penetration testing approach using openplc and factory io”. In: *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE. 2023, pp. 265–270.

[37] Mingeun Kim et al. “Firmae: Towards large-scale emulation of iot firmware for dynamic analysis”. In: *Proceedings of the 36th Annual Computer Security Applications Conference*. 2020, pp. 733–745.

[38] Giancarlo Fortino et al. “Enabling faster security assessment of re-hosted firmware”. In: *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. IEEE. 2022, pp. 1–6.

[39] Marcus Geiger et al. “An analysis of black energy 3, crashoverride, and trisis, three malware approaches targeting operational technology systems”. In: *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Vol. 1. IEEE. 2020, pp. 1537–1543.

[40] Pavel Kozak, Ivo Klaban, and Tomáš Šlajs. “Industroyer cyber-attacks on Ukraine’s critical infrastructure”. In: *2023 International Conference on Military Technologies (ICMT)*. IEEE. 2023, pp. 1–6.

[41] Juan C Olivares-Rojas et al. “Towards cybersecurity of the smart grid using digital twins”. In: *IEEE Internet Computing* 26.3 (2021), pp. 52–57.

[42] Kirti Gupta et al. “On the assessment of cyber risks and attack surfaces in a real-time co-simulation cybersecurity testbed for inverter-based microgrids”. In: *Energies* 14.16 (2021), p. 4941.

[43] Amardeep B Shitole et al. “Real-time digital twin of residential energy storage system for cyber-security study”. In: *2021 IEEE 2nd International Conference on Smart Technologies for Power, Energy and Control (STPEC)*. IEEE. 2021, pp. 1–6.

[44] Seerin Ahmad et al. “Advanced persistent threat (APT)-style attack modeling and testbed for power transformer diagnosis system in a substation”. In: *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE. 2022, pp. 1–5.

[45] Raffaele Cuorvo et al. “Securing industrial systems: A testbed for cyber-defense evaluation and data collection”. In: *2024 20th International Conference on Network and Service Management (CNSM)*. IEEE. 2024, pp. 1–7.

[46] Jamie Thorpe et al. *ADROC: An Emulation Experimentation Platform for Advancing Resilience of Control Systems*. Tech. rep. Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), 2022.

[47] Jamie Thorpe et al. “A Cyber-Physical Experimentation Platform for Resilience Analysis”. In: *Proceedings of the 2022 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*. Sat-CPS ’22. Baltimore, MD, USA: Association for Computing Machinery, 2022, pp. 3–12. ISBN: 9781450392297. DOI: 10.1145/3510547.3517916. URL: https://doi.org/10.1145/3510547.3517916.

[48] Immanuel Hacker et al. “A co-simulation environment to evaluate cyber resilience in active distribution grids utilising behind-the-meter assets”. In: *Electric Power Systems Research* 230 (2024), p. 110254.

[49] Muhammad M Roomi et al. “Auto-SGCR: Automated generation of smart grid cyber range using iec 61850 standard models”. In: *IEEE Open Journal of the Industrial Electronics Society* (2025).

[50] Thai-Thanh Nguyen, Rahul Kadavil, and Hossein Hooshyar. “A real-time cyber-physical simulation testbed for cybersecurity assessment of large-scale power systems”. In: *IEEE Transactions on Industry Applications* 60.6 (2024), pp. 8329–8340.

[51] Khandaker Akramul Haque et al. “Cyber-physical emulation and threat scenario simulation for enhanced microgrid resilience”. In: *IEEE access* (2025).

[52] Ömer Sen, Nathalie Bleser, and Andreas Ulbig. “Digital twin for evaluating detective countermeasures in smart grid cybersecurity”. In: *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE. 2023, pp. 1–6.

[53] Shuvangkar Chandra Das, Tuyen Vu, and Herbert Ginn. “Scalable cyber-physical testbed for cybersecurity evaluation of synchrophasors in power systems”. In: *IET Cyber-Physical Systems: Theory & Applications* 10.1 (2025), e12106.

[54] Jay Johnson et al. “Assessing DER network cybersecurity defences in a power-communication co-simulation environment”. In: *IET Cyber-Physical Systems: Theory & Applications* 5.3 (2020), pp. 274–282.

[55] Ioannis Zografopoulos et al. “Cyber-physical energy systems security: Threat modeling, risk assessment, resources, metrics, and case studies”. In: *IEEe Access* 9 (2021), pp. 29775–29818.

[56] Per-Arne Jørgensen et al. “Building a hardware-in-the-loop (HiL) digital energy station infrastructure for cyber operation resiliency testing”. In: *Proceedings of the 3rd international workshop on engineering and cybersecurity of critical systems*. 2022, pp. 9–16.

[57] Esteban Damián Gutiérrez Mlot, Jose Saldana, and Ricardo J Rodríguez. “Towards a testbed for critical industrial systems: Sunspec protocol on der systems as a case study”. In: *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE. 2022, pp. 1–4.

[58] Kalinath Katuri et al. “Experimental impact analysis of cyberattacks in power systems using digital real-time testbeds”. In: *2023 IEEE Belgrade PowerTech*. IEEE. 2023, pp. 1–6.

[59] Petr Blazek et al. “Smart grids transmission network testbed: Design, deployment, and beyond”. In: *IEEE Open Journal of the Communications Society* 6 (2024), pp. 51–76.

[60] Manolya Atalay and Pelin Angin. “A digital twins approach to smart grid security testing and standardization”. In: *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*. IEEE. 2020, pp. 435–440.

[61] Adam Bartusiak et al. “Extended gap analysis: An approach for security assessment of critical infrastructures”. In: *2022 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE. 2022, pp. 1–6.

[62] Adam Bartusiak et al. “First step into automation of security assessment of critical infrastructures”. In: *Sustainable Energy, Grids and Networks* 36 (2023), p. 101139.

[63] Daisuke Mashima et al. “Towards automated generation of smart grid cyber range for cybersecurity experiments and training”. In: *2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)*. IEEE. 2023, pp. 49–55.

[64] Ghada Elbez et al. “Insights and Lessons Learned from a Realistic Smart Grid Testbed for Cybersecurity Research”. In: *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. 2025, pp. 805–812.

[65] İsmail Erkek and Erdal Irmak. “Enhancing cybersecurity of a hydroelectric power plant using its digital twin model”. In: *2024 12th International Conference on Smart Grid (icSmartGrid)*. IEEE. 2024, pp. 372–377.

[66] Midhya Mathew and Faruk Kazi. “Hardware-in-Loop (HIL) testbed design of thermal power plant for threat modeling and attack vector analysis”. In: *International Journal of Critical Infrastructure Protection* 45 (2024), p. 100675.

[67] Yun Guo, Aijun Yan, and Junjie Wang. “Cyber security risk analysis of physical protection systems of nuclear power plants and research on the cyber security test platform using digital twin technology”. In: *2021 International Conference on Power System Technology (POWERCON)*. IEEE. 2021, pp. 1889–1892.

[68] RABE Silva et al. “Development of the asherah nuclear power plant simulator for cyber security assessment”. In: *Proceedings of the international conference on nuclear security, vienna, austria*. 2020, pp. 10–14.

[69] Andrew Stuart Hahn et al. *Automated Cyber Security Testing Platform for Industrial Control Systems.* Tech. rep. Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), 2021.

[70] Israel Barbosa De Brito and Rafael T de Sousa Jr. “Development of an open-source testbed based on the modbus protocol for cybersecurity analysis of nuclear power plants”. In: *Applied Sciences* 12.15 (2022), p. 7942.

[71] Filippo Rebecchi et al. “A digital twin for the 5G era: The SPIDER cyber range”. In: *2022 IEEE 23rd international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*. IEEE. 2022, pp. 567–572.

[72] Chiara Grasselli et al. “An industrial network digital twin for enhanced security of cyber-physical systems”. In: *2022 international symposium on networks, computers and communications (ISNCC)*. IEEE. 2022, pp. 1–7.

[73] Roberto Bruschi et al. “Managing 5G network slicing and edge computing with the MATILDA telecom layer platform”. In: *Computer Networks* 194 (2021), p. 108090.

[74] Antonio Pastor et al. “The Mouseworld, a security traffic analysis lab based on NFV/SDN”. In: *Proceedings of the 13th international conference on availability, reliability and security*. 2018, pp. 1–6.

[75] Danielle Dauphinais et al. “Automated vulnerability testing and detection digital twin framework for 5G systems”. In: *2023 IEEE 9th International Conference on Network Softwarization (NetSoft)*. IEEE. 2023, pp. 308–310.

[76] Simone Soderi, Daniele Masti, and Yuriy Zacchia Lun. “Railway cyber-security in the era of interconnected systems: a survey”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.7 (2023), pp. 6764–6779.

[77] Hongxue Chen et al. “Comparative Studies of Security Assessment Methods for Railway Control Systems”. In: *2024 IEEE 29th Pacific Rim International Symposium on Dependable Computing (PRDC)*. IEEE. 2024, pp. 201–207.

[78] Bradley Potteiger et al. “Simulation Testbed for Railway Infrastructure Security and Resilience Evaluation”. In: *NIST Publications* (2020).

[79] Chiara Grasselli et al. “A digital twin for enhanced cybersecurity in connected vehicles”. In: *2023 23rd International Conference on Transparent Optical Networks (ICTON)*. IEEE. 2023, pp. 1–4.

[80] Dongxian Shi et al. “A CAN Bus Security Testbed Framework for Automotive Cyber-Physical Systems”. In: *Wireless Communications and Mobile Computing* 2022.1 (2022), p. 7176194.

[81] Marek Wehmer et al. “NOTHORG; A digital twin for ECU security testing”. In: *Proceedings of the 2nd Cyber Security in CarS Workshop*. 2025, pp. 1–9.

[82] Stefan Marksteiner et al. “Using Cyber Digital Twins for Automated Automotive Cybersecurity Testing”. In: *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2021, pp. 123–128.

[83] European Union. *UN Regulation No. 155 — Uniform Provisions Concerning the Approval of Vehicles with Regards to Cyber Security and Cyber Security Management System*. Jan. 10, 2025. URL: https://eur-lex.europa.eu/eli/reg/2025/5/oj.

[84] Giovanni Paolo Sellitto et al. “A cyber security digital twin for critical infrastructure protection: The intelligent transport system use case”. In: *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer. 2021, pp. 230–244.

[85] Massimiliano Masi et al. “Securing critical infrastructures with a cybersecurity digital twin”. In: *Software and Systems Modeling* 22.2 (2023), pp. 689–707.

[86] Ahmed Amro and Vasileios Gkioulos. “Communication and cybersecurity testbed for autonomous passenger ship”. In: *European Symposium on Research in Computer Security*. Springer. 2021, pp. 5–22.

[87] Franck Sicard, Estelle Hotellier, and Julien Francq. “An industrial control system physical testbed for naval defense cybersecurity research”. In: *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2022, pp. 413–422.

[88] MITRE Corporation. *MITRE ATT&CK: Adversarial Tactics, Techniques, and Common Knowledge*. https://attack.mitre.org/. Accessed: 2025-10-28. 2025.

[89] Nickolaos Koroniotis et al. “The sair-iiot cyber testbed as a service: A novel cybertwins architecture in iiot-based smart airports”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.2 (2021), pp. 2368–2381.

[90] European Parliament and Council of the European Union. *Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union*. `https://eur-lex.europa.eu/eli/dir/2016/1148/oj`. Official Journal of the European Union, L 194, 19 July 2016, pp. 1–30. 2016.

[91] Marietheres Dietz et al. "Employing digital twins for security-by-design system testing". In: *Proceedings of the 2022 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*. 2022, pp. 97–106.

[92] Akhilesh Raj et al. "Autonomous Cybersecurity Testbed for Operational Technology Networks". In: *Proceedings of the 7th Workshop on Design Automation for CPS and IoT*. 2025, pp. 1–6.

[93] Ching-Huang Lin and Ci-You Liou. "Scenario-based IIoT Security Testbed: Implementation using the water pipeline control system scenario as an example". In: *2025 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*. IEEE. 2025, pp. 525–526.

[94] Andres Robles Durazno et al. "VNWTS: A virtual water chlorination process for cybersecurity analysis of industrial control systems". In: *2021 14th International Conference on Security of Information and Networks (SIN)*. Vol. 1. IEEE. 2021, pp. 1–7.

[95] Chris Churilla. "Conducting Attack Scenarios and Forensic Techniques in a Virtual ICS Test Bed". In: *2024 Annual Computer Security Applications Conference Workshops (ACSAC Workshops)*. IEEE. 2024, pp. 155–162.

[96] Sharad Agarwal, Awais Rashid, and Joseph Gardiner. "Old MacDonald had a smart farm: Building a testbed to study cybersecurity in smart dairy farming". In: *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*. 2022, pp. 1–9.

[97] Isabelle Brown-Cantrell and Thomas Morris. "Why the Chicken Crossed the Road: Commercial Egg Production Cybersecurity Threats and Testbed Design". In: *SoutheastCon 2025*. IEEE. 2025, pp. 856–861.

[98] Liuhuo Wan et al. "SATB: A testbed of IoT-based smart agriculture network for dataset generation". In: *International Conference on Advanced Data Mining and Applications*. Springer. 2022, pp. 131–145.

[99] Guillermo Francia and Gregory Hall. "Digital twins for industrial control systems security". In: *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE. 2021, pp. 801–805.

[100] Yuning Jiang et al. "Leveraging digital twin technology for enhanced cybersecurity in Cyber–Physical production systems". In: *Future Internet* 16.4 (2024), p. 134.

[101] Sheela Hariharan et al. "Towards a holistic approach to security validation of construction machinery through HIL systems". In: *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE. 2023, pp. 1–8.

[102] Sani M Abdullahi and Sanja Lazarova-Molnar. "Cybersecurity-oriented Digital Twins: A Double-Edged Sword or a Game Changer?" In: *2025 IEEE International Conference on Cyber Security and Resilience (CSR)*. IEEE. 2025, pp. 626–631.

[103] Andrew George Peck, Iain W Phillips, and Tim Watson. “Treasure Maps and Special Effects: The digital twin offering to threat actors”. In: *2024 IEEE Smart World Congress (SWC)*. IEEE. 2024, pp. 2175–2180.

[104] Antonio João Gonçalves de Azambuja et al. “Digital twins in industry 4.0–opportunities and challenges related to cyber security”. In: *Procedia Cirp* 121 (2024), pp. 25–30.

[105] Information System Authority (RIA). *Estonian Information Security Standard (E-ITS)*. Information System Authority. 2026. URL: https://www.ria.ee/en/cyber-security/management-state-information-security-measures (visited on 05/13/2026).

[106] Signal Foundation. *Signal Messenger*. 2026. URL: https://signal.org.

[107] WhatsApp LLC. *WhatsApp Security*. 2026. URL: https://www.whatsapp.com/security.

[108] Bundesamt für Sicherheit in der Informationstechnik (BSI). *IND.2.1 Allgemeine ICS-Komponente*. IT-Grundschutz Baustein. Accessed: 2026-05-01. BSI, 2023. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-GS-Kompendium_Einzel_PDFs_2023/08_IND_Industrielle_IT/IND_2_1_Allgemeine_ICS_Komponente_Edition_2023.pdf?__blob=publicationFile&v=3#download=1.

[109] Bundesamt für Sicherheit in der Informationstechnik (BSI). *SYS.1.5 Virtualisierung*. IT-Grundschutz Baustein. Accessed: 2026-05-01. BSI, 2023. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-GS-Kompendium_Einzel_PDFs_2023/07_SYS_IT_Systeme/SYS_1_5_Virtualisierung_Edition_2023.pdf?__blob=publicationFile&v=3#download=1.

[110] Bundesamt für Sicherheit in der Informationstechnik (BSI). *SYS.1.6 Containerisierung*. IT-Grundschutz Baustein. Accessed: 2026-05-01. BSI, 2023. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-GS-Kompendium_Einzel_PDFs_2023/07_SYS_IT_Systeme/SYS_1_6_Containerisierung_Edition_2023.pdf?__blob=publicationFile&v=4#download=1.

[111] *IEC 62443-2-1:2010. Industrial communication networks – Network and system security – Part 2-1: Establishing an industrial automation and control system security program*. International Electrotechnical Commission, 2010.

[112] *ISO 23247-1:2021. Automation systems and integration – Digital twin framework for manufacturing – Part 1: Overview and general principles*. International Organization for Standardization, 2021.

[113] NK Kandasamy et al. “An electric power digital twin for cyber security testing, research and education”. In: *Computers and Electrical . . .* (2022). Query date: 2025-04-06 16:06:15. URL: https://www.sciencedirect.com/science/article/pii/S0045790622003196.

[114] Matthias Eckhart and Andreas Ekelhart. “Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook”. In: Nov. 2019, pp. 383–412. ISBN: 978-3-030-25311-0. DOI: 10.1007/978-3-030-25312-7_14.

[115] Apurva Shet and Izzat Alsmadi. “An Empirical Analysis of Zero-Day Vulnerabilities Disclosed by the Zero Day Initiative”. In: *arXiv preprint arXiv:2512.15803* (2025). URL: https://arxiv.org/abs/2512.15803.

[116] Nicholas Carlini et al. *Assessing Claude Mythos Preview's Cybersecurity Capabilities*. Technical report/blog post published by Anthropic Red Team. Anthropic. Apr. 2026. URL: https://red.anthropic.com/2026/mythos-preview/ (visited on 05/07/2026).

[117] Bundesamt für Sicherheit in der Informationstechnik (BSI). *NIS-2 Kickoff – Das neue BSI-Gesetz und die regulierte Wirtschaft*. Webinar. Held on 20 January 2026, attended by the author; not publicly available. 2026.

[118] *Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011*. European Union, Dec. 27, 2022. URL: https://eur-lex.europa.eu/eli/reg/2022/2554/oj.

[119] Muna Al-Hawawreh and Elena Sitnikova. “Developing a security testbed for industrial internet of things”. In: *IEEE Internet of Things Journal* 8.7 (2020), pp. 5558–5573.

Appendix 1 – Non-Exclusive License for Reproduction and Publication of a Graduation Thesis[1]

I Kyrill Alexander Baron von Toll

1. Grant Tallinn University of Technology free license (non-exclusive license) for my thesis "Towards a Guideline for NIS2 Compliant Security Testing on Digital Twins of Cyber-physical Systems", supervised by Aleksi Oskar Johannes Kajander
    1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
    1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright
2. I am aware that the author also retains the rights specified in clause 1 of the nonexclusive license.
3. I confirm that granting the non-exclusive license does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act, or rights arising from other legislation.

28.05.2026

[1]The non-exclusive license is not valid during the validity of the access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 – Tables

Table 6. List of Simulation, Emulation, Virtualization, or other Software tools identified throughout the Thesis.

Begin of Table		
Tool name	Short Description	Sources using tool
AdvancedHMI	SCADA/HMI Emulation	[17], [6]
ANSYS	Physics Simulation	[17]
Asherah Nuclear Power Plant Simulator	Physics Simulation	[69], [70]
Bridgecommand	simulates navigational scenarios	[86]
C2WindTunnel	Physics Simulation	[17]
CANoe	simulates ECU	[101]
Cisco Modeling Labs (CML)	Network Simulation/Emulation	[76]
CoAPClient	Constrained application protocol communication	[119]
containernet	Network Emulation/Simulation	[52]
CORE	Network Simulation/Emulation	[76], [51], [53], [17], [55]
Docker	container virtualization	[12], [57], [52], [51], [28]
EasyBuilder	HMI Simulation	[35]
EasyModbusTCP	Modbus Communication Simulation	[91]
eMegasim	Physics Simulation	[58], [55]
ePhasorSim	Physics Simulation	[55]
ETAP eMTP	Physics Simulation	[55]
EVE-NG	Network Simulation/Emulation	[76]
EXata	network Simulation/ Emulation	[55]
ExataCPS	network Simulation/ Emulation	[58], [50]
EXSi 6.7	vm hypervisor	[119]
Factory I/O	3D Simulated Factories, including PLC Simulation	[95], [36]
GE iFix	SCADA/HMI Emulation	[17]
GEM5	Physics Simulation	[17]

Continuation of Table 6		
Tool name	Short Description	Sources using tool
GNS3	Network Simulation/Emulation	[76], [70], [86]
GRFICS	3D Simulater Factories, including PLC Simulation	[16], [92]
Gridlab-D	Physics Simulation	[55], [92]
HyperSim	Physics Simulation	[55]
ICSSIM	Python based ICS Simulation solution using Docker	[16]
IEC Server	field device Emulation	[16]
IEEE 30-bus Model	Physics Simulation	[17]
Instrument cluster simulator (ICSim)	CAN-Bus Simulation	[101]
InterPSS	Physics Simulation	[17]
LabVIEW	Physics Simulation	[17]
libIEC61850	C library for IEC61850 communication	[12], [59]
MatDyn	Physics Simulation	[17]
MATLAB Simulink	Physics Simulation	[6], [94], [68], [55], [43], [53], [17], [12]
MatPower	Physics Simulation	[17]
MiniMega	VM deployment	[69], [54]
Mininet	network Emulation/Simulation	[49], [76], [92], [28]
ModbusPal	Java-based Modbus communication Emulation	[16]
Modbus-tk	RTU Simulation	[6]
ModRSsim2	Modbus Communication Simulation	[6], [91], [70]
Mosaik 3.0	co-Simulation	[48], [52]
NetEm	Network Emulation	[55]
NETLAB+	virtual machine management	[26]
Netsim	IIoT Network & Device Emulator	[6]
NetToPLCsim	Interface between Network and PLC	[35]
Networksim	Network Simulation/Emulation	[17]
NMEASimulator	simulates navigational scenarios	[86]
Node-RED	SCADA , HMI,IoT, and PLC Simulation	[12], [45], [89], [119]
NS simulator	Network Simulation/Emulation	[17]
NS-2	Network Simulation	[55]
NS-3	Network Simulation/ Emulation	[6], [55], [59]
OMNeT++	Network Simulation/Emulation	[78], [16], [17]

Continuation of Table 6		
Tool name	Short Description	Sources using tool
OPAL-RT	Physics simulator	[58], [50], [53]
Open vSwitches	Network device Emulation	[45], [12]
OpenDSS	Physics Simulation	[55]
OpenModelica	Critical process simulator	[6]
OpenMUC	Physics Simulation	[17], [59]
OpenPLC	PLC Emulation/Simulation	[6], [36], [49], [17], [92], [95], [99], [119], [70], [91], [97]
OpenPLC Client	Python API emulating Modbus TCP API calls	[92]
OpenStack	Cloud computing infrastructure	[59]
OPNET	Network Simulation/Emulation	[17]
PacketSender	Network communication Simulation	[86]
Pandapower	Power-system Simulation	[48], [49], [52]
pfSense	simulates Network equipment	[36], [26]
PowerWorld	power-system Simulation	[51], [17]
Pydnp3	Python lib for dnp3 communication	[6]
PyModbus	python lib for modbus communications	[45]
pySynphasor	python lib for IEEE C37.118 communication	[53]
QEMU	IIoT Network & Device Emulator	[6]
QTester104	Field Device Communication Emulation	[16]
rettij	Network Emulation/Simulation	[52]
RevRun	Emulation experiment control and data collection	[46]
RINSE	Network Simulation/Emulation	[16], [17]
Rodbus	Rust-based Modbus communication Emulation	[16]
RTDS	Physics Simulation	[55], [58], [50]
S7-Plcsim	PLC Simulation based on S7	[35]
ScadaBR	SCADA/HMI Emulation	[6], [36], [36], [49], [91], [95], [70], [97]
SCADASim	SCADA/HMI Emulation	[16]
SCADAVT	SCADA/HMI Emulation	[16]
SCEPTRE	simulates large-scale networks	[69], [46], [54]
SecuriCAD	IT-infrastructure Simulation	[84], [100]

Continuation of Table 6		
Tool name	Short Description	Sources using tool
Siemens Modbus servers	modbus server	[69]
Siemens PLCSIM Advanced	PLC Simulation	[69]
SIMANTIC STEP 7	PLC Simulation/Emulation	[35]
SIMATIC S7-PLCSIM Advanced V3.0	PLC Emulation/Simulation	[94]
SimPy	Network Simulation	[55]
Sphere Hypervisor	virtual machine deployment	[26]
SUMO	Physics Simulation	[17], [78]
Tcpreplay	Network communication Simulation	[86]
Totally Integrated Automation (TIA) Portal	Siemens Framework to manage whole automation system	[35]
TrueTime	Physics Simulation	[17]
txThings	Python library for constrained application protocol	[119]
Typhoon HIL	Physics Simulation	[55]
UAVSim	Physics Simulation	[17]
Vagrant	virtual machine deployment	[53]
Veins	combination of SUMO and OMNeT++	[78]
Virtual Edge Device (VED)	edge device Simulation	[48]
virtual Ethernet (veth)	ethernet link Emulation	[28]
Virtualbox	VM hypervisor	[86], [95], [28], [12]
VMWare	VM hypervisor	[35], [17], [86], [119]
VTeststudio	simulates ECU	[101]
VyOS	simulates Network equipment	[70]
WebGME	metamodeling framework	[78]
WinPP104	Digital Substation Simulation	[61]
ZeroMQ	Network Simulation/Emulation	[75].
End of Table		

Appendix 3 – Questionnaire Text

Introduction and Consent

Thank you for participating in this qualitative survey on the use of Digital Twin technology for NIS2-compliant assessment of the effectiveness of risk management measures. This survey is conducted as part of a Master's Thesis by Kyrill Baron von Toll under the supervision of Dr. Aleksi Kajander at the Tallinn University of Technology. The Survey consists of six open-ended questions. Participation is voluntary and can be revoked at any point before submission of the thesis on May 20, 2026. Your responses will be used exclusively for the purposes of the thesis. It is possible to participate anonymously. For any further questions, please contact kybaro@taltech.ee.

Do you consent to participate in this Study?

Yes No

As this survey is conducted for a Master's Thesis at Tallinn Technical University, its contents and your responses to questions 1 to 6 will be made publicly available in the university's library. All responses represent the personal opinions of the participants and should not be interpreted as legal or regulatory advice.

Do you consent to your answers being publicised?

Yes No

If you wish, your personal information can be withheld. In this case, your answers would be attributed to an internal ID, and an abstraction of your profession/role would be provided.

Do you request your personally identifiable information to be anonymised?

Yes No

After the first stage of this survey is complete, interim results will be presented to a subset of participants for validation. This validation would be conducted as short interviews via video call. The validation stage is planned for the second half of March 2025.

Would you be willing to participate in an interview for validation of results?

Yes No

Please write down your full name.

Please describe your background and current position. Please describe your company/institution.

Survey Questions

Article 21 §2.f of the NIS2 directive requires "policies and procedures to assess the effectiveness of cybersecurity risk-management measures;" [2]

1. How would you fulfil this requirement in practice? Please provide examples.

Cyber-Physical Systems (CPS), also called Operational Technology (OT) systems, are systems where digital control causes physical effects and vice versa. Examples include automated factory lines, electronically controlled power plants or railway signal control networks. Such CPS commonly use hardware and software with longer life-cycles than other IT systems. Towards the end of their life cycle, these hardware and software components are often referred to as legacy systems.

2. In your experience, are there differences in how cybersecurity risk manage-

[2]DIRECTIVE (EU) 2022/2555. NIS 2 Directive. `https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng`

ment and its effectiveness should be assessed for CPS or legacy systems compared to other IT systems?
If so, please describe these differences and how they would affect your answer to Question 1.

A digital twin (DT) is a partially digital representation of an IT or OT system. In the context of this survey, we include terms such as "testbeds" and "cyber ranges". The purpose of a DT is to mimic its real counterpart as faithfully as possible through means such as simulation, virtualisation, or emulation of system components.

3. Could a DT be used to assess the effectiveness of cybersecurity risk-management measures? Does the type of DT, e.g., partly physical, virtualised, or simulated, impact this?

In DTs, the term fidelity describes how accurately the DT represents its real counterpart. At its simplest, one way to express fidelity is as a percentage. An example is the rate at which the DTs' and the real systems' responses to various inputs match.

4. What level or type of fidelity DT would be necessary for your answer in Question 3, or to change your answer to Question 3?

Building and maintaining a DT can be costly, especially for large, distributed systems.

5. Would using DTs to achieve NIS2 compliance be worth the cost? How would this approach need to compare to alternatives to be viable?

6. Please leave any additional comments here:

Epilogue

Thank you very much for your participation in this survey! The search for participants is still ongoing. If you have colleagues or acquaintances whose input you think would be valuable, and who might be willing to participate, please feel free to forward this survey to them or let us know in the textbox below. Finally, please save your answers and send the PDF back to kybaro@taltech.ee. Thank you again, and have a good day!

Appendix 4 – Questionnaire Responses

In this appendix, the participants' answers are presented. Line-breaks were removed for LaTeXformatting reasons. No further changes have been made. Empty answers are indicated.

Question 1

How would you fulfil this requirement in practice? Please provide examples.

Participant 1

In my current employment, we handle confidential data that must be legally retained. Therefore, we maintain both redundant backups and replicated storage. We periodically restore selected datasets as well as perform full restoration tests to verify data integrity and system recoverability. The results are documented and compared against defined recovery objectives (e.g., recovery time and data loss thresholds). Deviations trigger corrective actions, ensuring the effectiveness of our business continuity measures is continuously evaluated.

Participant 2

Our company is regularly audited, both internally and externally. The external audit is conducted by a qualified and endorsed auditor of the BSI (Bundesamt für Sicherheit in der Informationstechnik).

Participant 3

Document risk-management processes and methodology and how those changed over time. Document how risk assessment (significance of risks and risk acceptance / tolerance levels) is percieved over time. Work with external consultancy to explore potential weaknessses in current risk-management measures. Make sure that risk-management covers a wide-range

of potential threats – not just the ones directly related to cybersecurity – to cover unexpected blindspots. Continously evaluate existing risks and threats by crowd-sourcing potential vulnerabilities among people familiar with the organizations systems and processes (e.g. employees, customers. . .).

Participant 4

1. Independent External Validation: We are SOC 2 Type II audited and conduct annual penetration testing 2. Continuous Automated Monitoring: We have Kertos for compliance automation and Vulnerability scanning via Aikido.dev 3. Structured Internal Governance: Periodic Management Reviews, Internal Audit Procedures, Vendor Effectiveness Assessments

Participant 5: Dr. Phillip Empl

For technical controls, I would define measurable verification activities such as automated security tests, vulnerability scans, configuration reviews, and targeted checks (e.g., identifying orphaned identity accounts or inactive privileged users). The results of these tests would be documented, tracked over time, and linked to remediation processes to ensure continuous improvement. For organizational controls, I would implement structured reviews and validation mechanisms. This could include internal audits, short control effectiveness questionnaires for process owners, walkthroughs of procedures, and sampling-based evidence checks to verify that defined processes (e.g., incident handling, access approvals) are actually followed in practice.

Participant 6

Policies and procedures alone do not "assess the effectiveness of Cybersecurity risk-management measures" ! Having a practical and well maintained ISMS (Information Security Management System) is the best start. A strong Security organization within the company with enough budget, talent and executive management support to be effective are key! Risk appetite and maturity all play key factors!

Participant 7

Management-Review and good KPIs Specific tests Emergency Exercises

Question 2

In your experience, are there differences in how cybersecurity risk management and its effectiveness should be assessed for CPS or legacy systems compared to other IT systems? If so, please describe these differences and how they would affect your answer to Question 1.

Participant 1

Yes, CPS and legacy systems require a different assessment approach than traditional IT systems. Real active security testing like vulnerability scanning can disrupt the operational system which is normally not really feasible. Therefore you need indirect methods. Also the focus relies more on incident response that for example exploit attempts. Compared to my answer in Question 1, which relied on periodic restoration tests for data integrity, CPS environments require additional validation that the system remains safe and operational during cyber incidents. Like for example performing simulations in isolated enviroment

Participant 2

As our company is not involved in manufacturing we do not use CPS. However, in a suitable business environment, such systems could be useful to enhance the efficiency and manageability of processes. On the other hand, such systems enable data collection and guidance of subsystems. As such, CPS can become a target for unintended purposes and misuse. Their extended life cycle further exacerbate the associated risk.

Participant 3

Acknowledge that the risks associated with CPS might increase the longer their hard-/software's lifecycle lasts and increase scrutiny year by year given the increased potential of someone finding – and exploiting – vulnerabilities. Consider human security factor as an increased risk and carefully evaluate existing training opportunities and awareness measures for potential mistakes caused by conveninence.

Participant 4

As I have not worked with such systems, my experience is virtually none. But I do have an opinion, as I think, that for such legacy systems the risk management needs to be even harder, as these systems do not get regular sercurity updates etc. As we do not use such systems ourselfs (perks of being a start-up), I would say that I do not change my previous answers.

Participant 5: Dr. Phillip Empl

Industrial systems have long life cycles and rely on custom hardware, so risk management must address both software and physical components, including hardware dependencies (e.g., HBOM), not just SBOM. Many common security tools focus on software and are less effective for embedded or hardware-level risks. In CPS environments, patching is often impractical due to operational or certification constraints. Therefore, effectiveness is assessed more through compensating controls, segmentation, monitoring, and formal risk acceptance, since certain risks cannot be fully eliminated.

Participant 6

CPS / IoT and OT all contain the same amount of effort to secure, lifecycle and manage - the key differences are that most companies don't have a common strategy or oversight across the various disciplines - many times regular IT only manages Infrastructure and EUC (End User Computing) - but not Manufacturing or Production systems, SCADA, OT, etc. - which is handled separately from Engineering teams - that creates a "shadow IT" effect and lack of clarity who "owns" what as far as maintaining good security practices and following industry standards ! It all boils down to maturity levels and who ultimately is responsible and who is accountable for Cyberrisk!

Participant 7

The methods of a Risk Assessment might not differ but the impact of identified risks definitely does. Hence effectiveness should be tested more thoroughly. Appropriate procedures and systems should be put into place.

Question 3

Could a DT be used to assess the effectiveness of cybersecurity risk-management measures? Does the type of DT, e.g., partly physical, virtualised, or simulated, impact this?

Participant 1

Yes, depending on the measures, a Digital Twin allows us to safely prove that security controls actually reduce operational risk, especially in CPS environments where testing on production systems is unsafe or infeasible. For example, in my case we could use a Digital Twin to test restoration procedures under realistic attack scenarios. The type of Digital Twin also clearly impacts what can be demonstrated. Pure simulations can validate concepts and risk models, whereas full testbeds provide a higher level of assurance but are more costly

Participant 2

DT structures are essential to pilot intended cybersecurity measures and processes before they are implemented in a real life production environment. In this role they help to mitigate risks of unintended features of such measures. Virtualization and emulation provide some insight, but the simulation in a full scale test environment would be preferred.

Participant 3

Simulating possible attack vectors as closely as possible by using a digital twin could definitely enhance evaluating existing practices of cybersecurity riskmanagement. DT might be especially helpful for countering risks that organizations don't want to test in production. A partly physical DT might be more useful for evaluating CPS risk management given it's closer resemblance of the system.

Participant 4

Yes, I think it can be vital to understand and test the technical risks, but it is still limited, as the personal/social risks cannot be created in a DT. I'd say it depends on the system you are testing. In my case, as we are a Cloud bases SaaS start-up, it does not make sense to create a (partially) physical DT, as all the software and infrastructure is hosted by vendors.

Participant 5: Dr. Phillip Empl

A Digital Twin (DT) can significantly support risk management in CPS, particularly in the context of IEC 62443-3-2 risk assessments. CPS environments involve complex relationships between hardware, operating systems, firmware, and application software, which must be systematically identified and analyzed during threat and risk assessment.

Participant 6

YES - Using DT can be quite effective as a risk management tool - yet it requires a higher level of investment and maturity that many organizations simply have not achieved - To simulate or testbed real systems requires intimate knowledge of the systems and processes - and a dedicated Team to manage the DT processes and results - not every company has those extra resources at their disposal!

Participant 7

A DT is good while drafting measures for the system in terms of architecture and dependencies. Due to complexity of the real stack of an IT System the DTs value is limited here. The definition of a DT already shows the problem: it is not a true copy of the real system. That said, it is obvious that the type and design of a DT has major impact on the validity of statements about cybersecurity measures.

Question 4

What level or type of fidelity DT would be necessary for your answer in Question 3, or to change your answer to Question 3?

Participant 1

The required fidelity depends on what is being evaluated. For restoration and recovery tests, the Digital Twin would need accurate system architecture, configurations, dependencies, and realistic network behavior, but not detailed physical process simulation. Functional and network behavior must be realistic enough that attacks and recovery behave similarly to the real system. If the fidelity is too low, the twin would only support conceptual validation and could no longer reliably demonstrate effectiveness, which would change my answer to

Question 3.

Participant 2

The representation of fidelity in a rate of percentage does not appear to be an ideal number. In fact, the DT would be a setup which is as similar as possible to the real life environment. If not possible or not economically feasible the DT would include the main structural aspects of the real life environment which have to be previously defined and characterized.

Participant 3

Risk management evaluation should of course never rely on DT alone and the insights generated using DT should be evaluated to make sure the responses match what would have occured in production. That being said I believe a 70+% fidelity rate might be sufficient to create valuable indicators of strengths / weaknesses in existing risk management while also not producing too many weak spots / false possitives when responses don't match the real system.

Participant 4

I don't think my company needs a very high level of fidelity. What's more important is that our providers maintain a high level of fidelity, since we—and many others—rely on their infrastructure and security.

Participant 5: Dr. Phillip Empl

In our approach, we focus primarily on modeling assets and data flows between components. While we decompose an industrial machine into its main hardware and software elements, we typically treat externally sourced components from suppliers as defined black boxes. That means we assess their interfaces, data exchanges, trust boundaries, and functional role within the system, but do not analyze their internal structure.

Participant 6

DT Fidelity from low - medium or high really depends on the application and industry - Once again, the willingness to invest in what's necessary to achieve high fidelity often is an

obstacle to achieve the desired results. Also, build out a capable DT team and maintaining continuity are big challenges these day - especially around finding, attracting and retaining good talent . .

Participant 7

That truely depends on the risk appetite of the organization and the risk classification of the very system in question. The higher the better. Money is probably the limiting factor.

Question 5

Would using DTs to achieve NIS2 compliance be worth the cost? How would this approach need to compare to alternatives to be viable?

Participant 1

Using Digital Twins for NIS2 compliance can be worth the cost, especially for critical or complex systems where testing on production systems is unsafe or not possible. A DT allows repeatable and realistic testing of measures such as detection, segmentation, and recovery without disrupting operations. The approach would be viable if it provides safer and more reliable evidence than alternatives. However, for small or simple IT environments, as in my case, the cost may outweigh the benefit.

Participant 2

From my point of view, NIS is a relatively simple and crude approach to enhancing cybersecurity. At best, a DT could simulate a reporting process in case of an applicable security incident. However, I would recommend real life exercises instead of setting up a DT environment.

Participant 3

I belive DT would be worth the costfor protecting a system against cybersecurity risks if a high fidelity rate can be achieved with reasonable costs. There might be an argument for DT not being worth the cost “just” for achieving NIS2 compliance if regulators also accept weaker, less-costly risk management evaluation measures.

Participant 4

In our case, it makes little to no sense for us to maintain such a system on a permanent basis. In addition to our production system, we have QA and staging systems that at least verify our own changes, such as those made to the codebase.

Participant 5: Dr. Phillip Empl

With the EU Cyber Resilience Act (CRA), machine builders and other manufacturers will increasingly be required to provide Bills of Materials (BOMs) for their products. This creates an opportunity for operators subject to NIS2, such as a brewery or other industrial facility, to integrate these BOMs into their own risk management processes. A practical approach would be to aggregate the BOMs of deployed machines, map them onto the existing network architecture (zones and conduits), and correlate this information with monitoring data in the SOC. This would enable visibility into affected assets when new vulnerabilities are disclosed and support risk-based prioritization of mitigation measures.

Participant 6

Using DTs to achieve NIS2 compliance is NOT worth the cost or effort. It's also the wrong trigger or reason to perform DTs in general - they are not solely meant for compliance purposes, although the results could be used in conjunction with said compliance. Given limited budgets these days - money is better spent on other areas of achieving compliance and maintaining it. NIS2 is only one of many other areas of compliance like DORA, TiSAX, CRA, etc. At the end of the day - ISO27001 is the guiding light!

Participant 7

This has to be answered by risk assessment for the system. If impact and damage is high, one is willing to spend more money. Second criteria could be whether a DT is the only possible way to work on that topic. There will be scenarios where a real-life copy or test system is cheaper and better.

Question 6

Please leave any additional comments here:

Participant 1

[No answer given]

Participant 2

[No answer given]

Participant 3

[No answer given]

Participant 4

[No answer given]

Participant 5: Dr. Phillip Empl

[No answer given]

Participant 6

Difficult to be more specific with such questions - as the complexity and variety of Cybersecurity "compliance" is dependent on industry, maturity and governance

Participant 7

A field of high complexity and variability: Thorough risk assessment upfront to any decision of building a DT or not seems to be a good advice to not loose lots of money. Along with the rise of AI the use of DTs should also be re-evaluated. Things may change dramatically here.