2024 IEEE/ACM 4th International Workshop on Engineering and Cybersecurity of Critical Systems and 2024 IEEE/ACM Second International Workshop on Software Vulnerability (EnCyCriS/SVM)

Interplay of Digital Twins and Cyber Deception: Unraveling Paths for Technological Advancements

Jessica Heluany, Ahmed Amro, Vasileios Gkioulos, Sokratis Katsikas
jessica.b.heluany@ntnu.no,ahmed.amro@ntnu.no,vasileios.gkioulos@ntnu.no,sokratis.katsikas@ntnu.no
Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology
Gjøvik, Norway

ABSTRACT

This research delves into the consolidation of Digital Twin and cyber deception technologies and explores their potential synergy for advancing cybersecurity processes. The study begins with a literature survey and market analysis, revealing a scarcity of mature scientific and commercial contributions in this domain. Most discussions remain theoretical, emphasizing the need for further research to address challenges and practically apply these technologies. Promising applications encompass cyber deception, anomaly detection, and threat intelligence, predominantly utilizing digital twin-based honeypots.

The paper contributes by proposing a high-level deception framework tailored for Operational Technology (OT) systems, with seven pivotal functions for a deception network, emphasizing the replication of realistic systems, attracting attackers, controlling connections, monitoring activities, and analyzing detected events. Moreover, an evaluation via a SWOT analysis highlights various strengths, weaknesses, threats, and opportunities inherent in this framework identifying potentially innovative directions such as applications of digital twins, and artificial intelligence. Strengths include improved defender control and enhanced security analysis, while challenges revolve around achieving high realism in digital twins and managing restoration complexities. This study sets a roadmap for further exploration into the effective integration of Digital Twin and honeypot technologies in cybersecurity contexts.

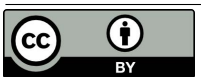
CCS CONCEPTS

• **Security and privacy** → **Systems security**; **Network security**.

KEYWORDS

digital twin, deception, honeypot, cybersecurity

ACM Reference Format:
Jessica Heluany, Ahmed Amro, Vasileios Gkioulos, Sokratis Katsikas. 2024. Interplay of Digital Twins and Cyber Deception: Unraveling Paths for Technological Advancements. In *2024 ACM/IEEE 4th International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS) and 2024 IEEE/ACM Second International Workshop on Software Vulnerability (EnCyCriS/SVM '24), April 15, 2024, Lisbon, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3643662.3643955



EnCyCriS/SVM '24, April 15, 2024, Lisbon, Portugal

ACM ISBN 979-8-4007-0565-6/24/04
https://doi.org/10.1145/3643662.3643955

1 INTRODUCTION

We are amidst the digitalization era across many fields. As computers and networks evolve, the power of AI, sensors, and ubiquitous connectivity is being harnessed to create advanced and complex digital systems. The Digital Twin concept emerged in response to the need to use virtual models to replicate real-world assets for advanced simulations and insights. The Digital Twin technology is a rapidly advancing technology in Industry 4.0 [30].

The paradigm of digital systems involves not only their core control/automation function in the industrial systems but also how they might expand the attack surface; this applies to digital twins because of data and information synchronization regarding the real assets, mostly in real-time [39]. We have observed interrelations between cybersecurity and digital twins in two fields: the security of the digital twin itself, and the use of a digital twin to support cybersecurity measures. Among the discussed areas regarding the use of digital twins to support cybersecurity is *cyber deception*, defined by NIST as "a system (e.g., a web server) or system resource (e.g., a file on a server) that is designed to be attractive to potential crackers and intruders like honey is attractive to bears." [15]

One objective of this paper is to capture the state of the art regarding the intersection between the digital twin and cyber deception technologies to identify and assess the opportunities for innovative research in this area. To achieve this, a literature survey and market desk research have been conducted and summarized in this paper. Moreover, based on the studied literature and identified market offerings, a high-level framework is proposed that captures the core functions that the interplay of these two technologies can contribute to the improvement of cybersecurity. The proposed framework is evaluated using a SWOT (strengths, weaknesses, opportunities, and threats) analysis. The analysis showcases the advantages and disadvantages of such a framework for supporting cybersecurity processes and assisted in identifying directions for future research.

The remaining of this paper is structured as follows: Section 2 provides a brief overview of digital twins and honeypots. Section 3 presents the findings of the literature review and of the market desk research on the intersection between the digital twin and cyber deception technologies. Section 4 presents our proposal for a conceptual framework for leveraging combinations of digital twins and honeypots targeting integrated Information Technology (IT) and Operational Technology (OT) systems, whilst the results of

the evaluation of the framework by means of SWOT analysis are discussed in Section 5. Section 6 discusses possible future research pathways as well as limitations of the present work. Finally, Section 7 summarizes our conclusions.

2 BACKGROUND AND RELATED WORK

This section provides a brief overview of digital twins and honeypots, focused on the concepts that are more relevant to our study.

2.1 Digital Twins

Several digital twin frameworks have been proposed, differing from the concept definition to the number of layers and considerations of lifecycle management. One definition, among many others, is from the Internet Engineering Task Force (IETF): "a virtual instance of a physical system (twin) that is continually updated with the latter's performance, maintenance, and health status data throughout the physical system's life cycle." [39]. Similarly, the Digital Twin Consortium defines it as "a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity" [9]. The aspects of synchronization frequency, and simulation fidelity can vary according to the needs of each use case, and even result in sub-terminologies, as suggested by Kritzinger [21] when defining *Digital Model*, *Digital Shadow*, and *Digital Twin* based on the level of data integration. In their terminology, the data synchronization from the physical object to the virtual object is done manually for the digital model, while for digital shadow it is manual from the digital to the physical object and automatic in the opposite direction. Lastly, the data flow in a digital twin would be automatic in both directions. Following this terminology, actually what is observed in most digital twin studies would be named a digital model, highlighting the simulation aspect provided by this conceptual technology; this is also the approach we follow in this work.

2.2 Honeypots

The purpose of this deception technique is to simulate systems, services, or data, making them attractive to attackers. Lance Spitzner, founder of the non-profit organization 'The Honeynet Project' defined honeypots as "a security resource whose value lies in being probed, attacked, or compromised" [32]. While some scholars call *honeynet* a network of honeypots, most utilize honeypots regardless of the number of devices being mimicked. Among common use cases, honeypots can be utilized to divert attackers or collect information on their methods, thus providing resources for analysis and potential improvements of the security measures in the corresponding real system or similar ones.

Honeypot classification is usually based on features such as purpose, role, application, deployment nature (virtual, physical, or hybrid), scalability, and level of interaction. Given the scope of this work, the level of interaction is the most relevant feature; this is mostly measured based on the complexity and extent of available request and response activities and ranges from low to high interaction honeypots. While a low-interaction honeypot provides limited and pre-designed functions with no access to the operating system, a high-interaction level honeypot mimics several functions and gives access to the operating system. Similar to the simulation fidelity, the interaction level of honeypots depends on the use case. Usually, low-interaction honeypots are utilized as Intrusion Detection/Prevention Systems (IDS/IPS), while high-interaction applications focus on collecting information for studying and analyzing the attackers' methods in more complex attack scenarios, such as in Advanced Persistent Attacks (APTs).

2.3 MITRE Engage

The MITRE Corporation has developed a framework for planning and discussing adversary engagement operations, called *Engage* [4]. The framework divides the engagement operations into three phases, namely *Prepare, Operate,* and *Understand.* The *Prepare* phase focuses on preparing the environment and on the information required to tailor the environment for the targeted attackers. The *Operate* phase focuses on establishing capabilities or tactics, namely *Expose, Affect,* and *Elicit.* Many of these capabilities involve cyber deception actions, such as luring the attackers and making the environment appear realistic and lively. Lastly, the *Understand* phase focuses on after-action activities such as analyzing and understanding the adversarial behaviors. While it does not refer particularly to Operational Technology or digital twins, the framework refers to a variety of techniques or methods for cyber deception or the utilization of simulated aspects of the environment that are highly relevant to the framework proposed in this paper.

3 INTERSECTION OF DIGITAL TWINS AND CYBER DECEPTION

In this section, the findings of the literature review and of the market desk research on the intersection between the digital twin and cyber deception technologies are presented. The former capture the academic perspectives on the topic, whereas the latter capture the current relevant market offerings.

3.1 Academic Perspective

A comprehensive literature survey was conducted to capture the state-of-the-art concerning the intersection between digital twins and honeypots. The databases utilized were Google Scholar and Scopus due to their multiple input sources, having as keywords the combination of "digital twin" AND ("honeypot" OR "deception") in their titles or abstracts. As inclusion criteria, only works that in fact approached a combination of the technologies were selected. The volume of identified literature is relatively small; this highlights the scarcity of scientific contributions in the field. In total, 15 works were discovered discussing the two topics in conjunction with each other. Additionally, 6 other works refer to the concepts without specifically utilizing the "digital twin" or the "honeypot" terminologies; rather, these works describe advanced system modelling and simulation within the context of cyber deception and trapping.

The level of maturity in the works is generally low. Most articles only mention or provide theoretical discussion or analysis regarding the compatibility of the two technologies. Some works have proposed frameworks for integrating both concepts through architectures, models, tools, and methods [28, 37], and few have reached an implementation level or prototyping [11, 18, 25]. This

Table 1: Summary of the patterns of intersections between the digital twin and honeypot technologies

Pattern	Brief work description
DT of a system functioning as a HP	- DT-Based Honeypots as a possible application of DT for cybersecurity [13, 14, 16] - Building a DT of an IoT device and using it as a HP. [25] - A DT of a system can perform a function such as a HP [23] - A DT prototypes as HP for Security-By-Isolation (SBI) gateway [11] - A DT behaves as the real system and functions as a HP. Underneath, it is implemented based on Large Language Models (LLM) to provide fake, yet realistic responses to attackers' commands. [24] - Counteracting Information Threats Using HP Systems Based on a Graph of Potential Attacks [36] - IOT honeynet for military deception and indications and warnings [18] - Towards systematic honeytoken fingerprinting [33]
DT of a HP to extend the HP functionality	- Creating a DT of a systems HP for expanded cybersecurity functionality such as anomaly detection and threat intelligence. [28] - Classifying resilience approaches for protecting smart grids against cyber threats [34]
HP as a security solution for a DT	- Discussing HPs as a general security solution for DTs [7]
DT and HP utilized separately	- DT as a general direction for enhancing cyber security. While HPs among the critical defense approaches. In smart grid [38] and manufacturing [27].
HP supports DT	- DT utilized for security attack simulation in SOCs based on the threat intelligence data received from HPs. This makes HP function as a data source for DT-enhanced SOC. [13, 16]

DT: Digital Twins ; HP: Honeypots.

calls for further investigation in the field to address the raised challenges, evaluate the proposed concepts, and assess the utility of the applications. The observed applications include cyber deception [7, 11, 13, 14, 16, 18, 23–25, 33, 36], anomaly detection [28], threat intelligence [13, 16, 28], security simulation [13, 16], and as a general direction for cybersecurity and resilience [27, 34, 38].

The observed patterns of intersection between the technologies can be categorized into five groups. We summarize the patterns and brief descriptions of the works in Table 1. It can be observed that the most common pattern is using digital twins as honeypots. While many works highlight the opportunities of combining the technologies, some are raising a profound challenge. The main discussed challenge is the amount of information that high-fidelity digital twins might provide for the attackers, thus facilitating the crafting of sophisticated targeted attacks [14, 19]. This suggests an interesting research direction, namely to investigate the optimal fidelity level of a digital twin-based honeypot so as to be realistic enough without fully revealing the system structure. Another suggested future direction is integrating digital twins into cyber ranges. This helps create an accurate representation of the physical environment as a honeypot. Consequently, learning more about digital twin-based honeypots [16]. On the other hand, honeypots within the context of cyber ranges can become data sources. The recorded adversarial behaviour extended from interacting with the honeypots in the cyber range environments can be forwarded for simulation within the digital twins [13].

3.2 Market Offerings

Investigating market offerings on the intersection between the digital twin and honeypot technologies has led to the identification of some products and tools involving the two technologies. Our investigation is only limited to searching resources on the web; no tools or products were tested. The observed applications of the offerings span across cyber ranges, cyber deception, data loss detection, and emergency preparedness. A common pattern of intersection is similar to that identified in the academic literature, namely digital twin-based honeypots. We have identified several solutions for cyber deception and data loss detection. A summary of such solutions follows:

- DarkStax™ [12]: This platform uses digital twins to create realistic environments for attackers while deterring them from the real environment. Honeypots are used for cyber deception.
- ActiveBehavior™ [10]: This solution enriches deception environments, such as honeypots, with a lively appearance to convince attackers that they are in live systems. Digital twins are used to create operationally realistic environments for protection.
- Honey Trace [2]: This solution uses honeypots for data loss detection. Digital twins are used to create twins of "documents" to embed credentials and catch anyone who tries to use them.
- Shadow Figment [5]: This solution implements interactive decoys (honeypots) mapped to simulations of real OT systems (digital twins) for control systems.
- Simulation and Verification Platform for Industrial Control Information Security in Water Industry [3]: This service provides a simulation environment for water management digital twins to enhance emergency response capabilities through honeypot deployment.
- SECTRIO [6]: This solution uses deception technologies to create digital twins of OT assets to lure attackers away from real ones.

- ThreatWise™ [1]: This solution employs threat sensors to engage attackers with a web service-like architecture mimicking real environments (digital twins), keeping attackers away from real assets while revealing their adversarial techniques.

4 CYBER DECEPTION FRAMEWORK FOR OPERATIONAL TECHNOLOGY ENABLED BY DIGITAL TWIN

The studied literature and observed market offerings motivated us to propose a conceptual framework targeting the combination of digital twin and honeypot technologies for integrated IT and OT systems. The proposed framework is based on NIST Cyber Cybersecurity Framework (NIST CSF) [29] and the MITRE Engage framework discussed in section 2.3, aiming to combine the core functions of NIST CSF that focus on cybersecurity management, with the core functions of MITRE Engage, which explore deception. The new release of NIST CSF consists of six core functions: Govern, Identify, Protect, Detect, Respond, and Recover. On the other hand, targetting the engagement of attackers, the MITRE Engage core functions are: Collect, Detect, Prevent, Direct, Disrupt, Reassure, and Motivate.

The proposed deceptive defence framework combines both NIST and MITRE functions, being composed of seven core functions as shown in Figure 1, namely creating realistic replicas (Mimic); attracting attackers with decoys (Attract); deploying and orchestrating replicas (Deploy); intercepting and controlling connections with attackers (Engage); monitoring and logging attacker activities (Monitor); resetting the network to its original state (Reset); and analysing and responding to detected activities (Analyse). The suggested framework was subsequently utilized to explore venues of innovation and impactful research directions.

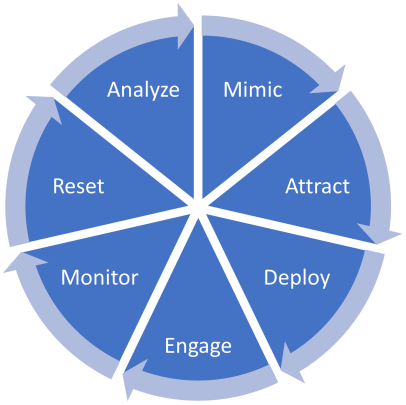


Figure 1: The core functions of the proposed framework.

Mimic: The mimicking function focuses on creating representations (replicas) of systems' components and their network with the possibility of continuous updates from the real assets and with different fidelity levels.

The main purpose of this function is to increase the degree of realism of the system replicas. This can achieve several objectives. One objective is to increase the spectrum of engaging aspects in the replicas for the attackers and increase their engagement time during the engagement phase. Another objective is to provide the defenders with realistic system features to assist their tasks during the monitoring and analysis phases.

Several works have been observed that focus on the development of honeypots which mimic or simulate OT components such as PLCs (Programmable Logic Controllers), HMI (Human-Machine Interface), SCADA server (Supervisory control and data acquisition), and RTU (Remote Terminal Unit)[8, 22]. Among the surveyed works, this function was observed in creating digital twins of IoT devices [18, 25], network gateway [11], any device with Windows, Linux, or Macintosh operating system [24], and documents [2].

The digital twin technology would play a pivotal role in this function by creating replicas of systems with various degrees of fidelity, aiming to avoid revealing sensitive information to attackers.

Attract: The attraction function aims at making the replicas of components attractive for attackers to target and exploit through implanted decoys.

The main purpose of this function is twofold: to attract the attackers to engage with the replicas by advertising decoys on one hand, and to increase their engagement time on the other, by reflecting the perception of lively networks and systems hiding signs of simulation and deception that could make attackers realize that it is a honeypot rather than a real system that they are interacting with.

Among the decoys that could enrich the attract function are files (credentials or configurations), interactive services (SSH, Modbus, HTTP), configurations (disable access controls, firewall, no encryption), unpatched software versions, poor configuration of firewall rules, use of vulnerable ICS protocols, and lack of system hardening.

An important aspect to discuss in this function is the need for countering anti-honeypot techniques such as system-level fingerprinting [35], MAC fingerprinting [20], and time analysis [26]. The developers or operators of the replicas (i.e. digital twins) would need to consider this issue during development or configuration.

Deploy: This function focuses on deploying and orchestrating the replicas in the network.

Some works have been observed to plant a group of honeypots within the network to devise opportunities for attackers to interact with them instead of the real assets [3, 5]. Some works proposed what can be described as gradual deployment, that is, deploying honeypots based on potential attacks identified through analyzing the attackers' observed techniques [36].

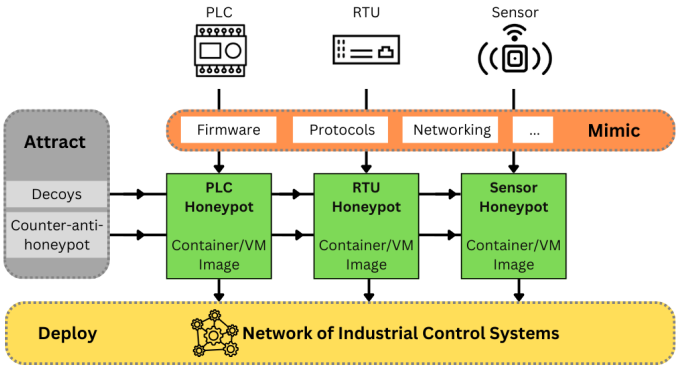


Figure 2: An example of the interplay between the mimic, attract, and deploy function

To exemplify how the mimic, attract, and deploy functions would work together, Figure 2 shows a topology where a PLC, an RTU and a sensor are mimicked through firmware, network protocols, and field measurements respectively. An individual honeypot image of each of these components would be mimicked, on which the attract function applies the relevant decoys and counter-anti-honeypot techniques before the VM images are deployed.

Engage: This function focuses on monitoring, intercepting, and controlling the engagement with attackers.

This function is tightly coupled with the mimicking functions, aiming mainly to provide realistic responses for the attacker and prolong their engagement to increase the chances of detection and the understanding of the attacker's intent and behaviour.

Among the observed approaches for the engagement function are the application of REST API [25], simple client-based communications API [11], or reverse proxies [17]. An innovative approach which we would like to highlight due to its novelty is the application of Large Language Models (LLMs), particularly ChatGPT, for creating digital twins of systems to operate as honeypots and engaging attackers as a Chatbot [24].

An important topic to highlight is that, despite the level of interaction, a honeypot is designed with a specific goal. In this sense, it is important to not consider high-interaction level honeypots as always better than low-interaction ones, because they might be designed to serve different purposes. Usually, low interaction honeypots can be a better solution for IDS [31], while high interaction honeypot helps in studying how attackers are behaving, and what vulnerabilities they are exploring, so that the system owner can improve the security measures of the real system.

In this paper, we stress the need for investigation of LLM-driven components for the engage function. Such direction holds a high prospect for innovation. Only a proof of concept exists by Mckee [24], but it does not focus on specialized OT assets. The concept of an LLM-driven engagement function is depicted in Figure 3. Specialized models of the OT assets could be developed and trained based on request-response pairs. Later, those models may be utilized for safely engaging attackers, as attackers will not be dealing with safety-critical systems but will still receive highly realistic responses. Additionally, such a solution is expected to exhibit low operational costs. A Proof of Concept (PoC) of such functionality is provided in Appendix A to inspire future work in this direction. Furthermore, using links to real assets, the model can update itself to maintain a realistic state. However, the latter option infers risks of an attack to propagate to the real asset.

Monitor: This function focuses on collecting logs, received from the replicas, reporting all the received and executed commands and sending them to be investigated in the analysis function.

The main aim of this function is to capture the attacker's activities in the network.

This function is relatively simple and several solutions exist that implement it. These include APIs [11], and Elasticsearch agents [13].

Reset: This function enables the restoration of the replicas to the original state. This is relevant to both honeypots and digital twins used by defenders.

The main aim of this function is to enhance the efficiency of the lifecycle i.e., shorter re-deployment time, of replicas, when an attacker leaves the network or the defenders use the replica for malware detonation for instance. This function enables the replicas to be re-used with less overhead.

This function is not commonly discussed in the observed literature, except for the LLM-based honeypot [24], due to the need for chat reset because of the limited context of existing LLMs.

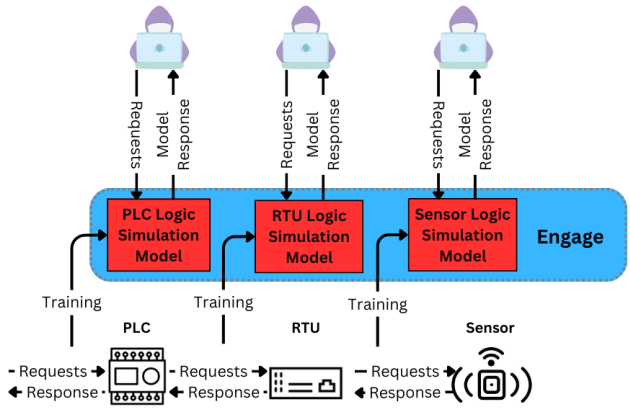


Figure 3: Conceptual overview of the LLM-driven engagement function

One possible direction forward could be the creation of snapshots of the components' digital twins to capture their safe states and utilize the snapshots in every usage cycle. Such a solution comes of course with the burden of managing a large number of snapshots in large networks.

Analyze: This function focuses on the analysis of and response to the captured adversarial activities.

The main aim of this function is to understand the adversarial behavior in the network to better defend against cyber threats.

This functionality of the interaction between digital twins and deception technology is claimed to benefit Security Operation Centers (SOCs) [13] and cyber ranges [1].

The solutions that could deliver this functionality include a group of analysis toolsets for a variety of cybersecurity analysis functions.

The utilization of high-fidelity replicas for supporting the analysis function is expected to open several novel research directions and enable technology innovations.

5 EVALUATION OF THE PROPOSED FRAMEWORK: A SWOT ANALYSIS

A critical evaluation of the proposed framework was conducted in an attempt to provide insights into the reasons why there is a scarcity of academic papers and industrial solutions that explore this joint approach. Each function was evaluated by applying a SWOT (Strengths- Weaknesses -Opportunities-Threats) analysis. The goal of such an analysis is to obtain a comprehensive view of the main concerns about each function, providing information for decision-making and researchers. A summary of the results of the analysis is provided in Table 2.

6 DISCUSSION

In this section, we will discuss the main aspects of our findings, to inform future research and development concerning the intersection between the digital twin and deception technologies. Additionally, we discuss the limitations of the present study.

6.1 Future Directions

Future research and development directions, as identified through the SWOT analysis, encompass several key areas. Firstly, enhancing the fidelity and realism of digital twins could be targeted, focusing on addressing challenges related to high fidelity for various components. A Multidisciplinary Competence Framework could also be developed to effectively establish and manage domain-specific competencies necessary for the comprehensive deception framework. Efficient snapshot management techniques for comprehensive restoration, particularly across a wide range of OT components, would be an important aspect. In bandwidth-constrained locations, exploration of bandwidth-adaptive logging solutions would support comprehensive logging without sacrificing network performance. Balancing resource utilization and attacker engagement, considering factors like delays in gradual deployment, is essential. Efforts could be made to mitigate honeypot detection, ensure resilience in resource-constrained areas, encourage collaborative configuration standardization, and enhance the resilience to zero-day exploits of the deception network itself . Additionally, measures should be taken to improve reporting accuracy; develop advanced threat data analytics; explore gradual deployment strategies; and support standardization efforts for digital twin technologies to ensure interoperability and consistency across implementations. These strategic directions are proposed to guide future efforts.

6.2 Limitations

Among the limitations of the conducted survey is the quality of considered works. Some of these works (master thesis and preprints) have not been peer-reviewed. Further, the market offerings were not tested, only analyzed based on their descriptions. The decision to include them was made so as to capture the trends for the intersections from comprehensive sources and assess the maturity level of the existing solutions rather than their quality.

7 CONCLUSIONS

The trend of digital twins is gaining significant attention in both academia and industry, as it has the potential to bolster the digital transformation of industry. This has led to an investigation into the usefulness of this trend for supporting cybersecurity processes. Our research has studied the intersection between digital twin technology and cyber deception to explore potential benefits for various industries and academic research directions. To achieve this, we conducted a comprehensive literature survey and market desk research, to capture the state-of-the-art applications of both fields side-by-side in academic works and industrial offerings.

The integration of digital twins and cyber deception is still an emerging direction with a limited number of mature works. While some works propose frameworks, models, and tools for integrating both concepts, few have reached the implementation level. Our findings suggest that digital twins can be used as honeypots in cyber deception, anomaly detection, threat intelligence, and security simulation. However, high-fidelity digital twins might provide too much information to attackers, allowing them to craft sophisticated targeted attacks. Future research should investigate the optimal fidelity level of digital twin-based honeypots and integrate them into cyber ranges to create an accurate representation of the physical environment.

Based on our findings, we proposed a cyber deception framework that utilizes aspects of digital twin technology. The framework includes creating realistic replicas (mimic), attracting attackers with decoys (attract), deploying and orchestrating replicas (deploy), intercepting and controlling connections with attackers (engage), monitoring and logging attacker activities (monitor), resetting the network to its original state (reset), and analyzing and responding to detected activities (analyze). These capabilities make the framework a comprehensive and robust security solution. The development of the framework has also led to the identification of novel components and research directions with innovation potential, including the use of Large Language Models for creating deceptive digital twins to engage attackers in a safe manner.

We evaluated the proposed framework using SWOT analysis and found that it is a great enabler for a group of cybersecurity processes, giving defenders increased control over attackers. However, we also highlighted some risks, such as increasing the attack surface, and we identified the specialized knowledge required for operating such a framework as a weakness.

Possible future directions of research and development include enhancing the fidelity and realism of digital twins, developing a multidisciplinary competence framework, efficient restoration techniques, and balancing resource utilization and attacker engagement. We also recommend improving the reporting accuracy, developing advanced threat data analytics, and supporting standardization efforts for digital twin technologies. Another direction could be the investigation of computational power required for running Large Language Models (LLMs) as honeypots on different platforms (e.g. personal computers).

We believe that our findings can inspire and motivate academics and industrial stakeholders interested in digital twins or cyber deception to consider integrating the functionality proposed in the

Table 2: Summary of the SWOT analysis of the proposed framework

Strengths	Weaknesses
- Improved defender control over attackers. - Increased effort and time for attackers, enabling better response. - Flexible with virtualization for adding components. - Early development opportunities, reducing future costs. - Low implementation costs, leveraging existing tech. - Enhanced security analysis with digital twins. - Risk reduction by containing attackers. - Release of critical resources via digital twins. - Quick reuse and learning with honeypots.	- Challenges in achieving high realism in digital twins. - Need for diverse domain-specific skills. - Management complexity of restoration snapshots. - Logging issues in bandwidth-constrained locations. - Storage and analysis challenges with comprehensive logs. - Unclear replication methods for some components. - Balancing resource utilization and attacker engagement. - Unclear triggers for restoration.
Threats	**Opportunities**
- Potential for attackers to deduce sensitive info due to high fidelity. - Risk of honeypot detection by attackers. - Resource consumption by excessive attacks. - Limited partner collaboration on digital twin configurations. - Risk of zero-day exploits bypassing engagement functions. - Vulnerabilities in the deception network used as an attack vector. - Risk of inaccurate reporting. - Limited data analytics for advanced threats.	- Research on gradual deployment and optimal resource optimization. - Identifying optimal fidelity levels for components. - Multidisciplinary collaboration among partners to improve industry offerings. - Support for standardization of digital twin technologies.

framework and the findings of the SWOT analysis in developing more robust and useful solutions.

REFERENCES

[1] [n. d.]. Cyber deception. https://metallic.io/knowledge-center/glossary/cyber-deception

[2] [n. d.]. HoneyTrace: Data loss detection. https://honeytrace.io/

[3] [n. d.]. Intelligent Water Digital Twin Security Simulation and Verification Laboratory. https://www.ticpsh.com/en/security/fzyz/swhy

[4] [n. d.]. MITRE EngageTM. https://engage.mitre.org/matrix/

[5] [n. d.]. Shadow Figment: Model-Driven Cyber Defense for Control Systems. https://www.pnnl.gov/available-technologies/shadow-figment-model-driven-cyber-defense-control-systems

[6] 2023. Top 7 Benefits of Deception Technologies in OT Security. https://sectrio.com/web-stories/top-7-benefits-of-deception-technologies-in-ot-security/

[7] Cristina Alcaraz and Javier Lopez. 2022. Digital twin: A comprehensive survey of security threats. *IEEE Communications Surveys & Tutorials* 24, 3 (2022), 1475–1503.

[8] Daniele Antonioli and Nils Ole Tippenhauer. 2015. Minicps: A toolkit for security research on cps networks. In *Proceedings of the First ACM workshop on cyber-physical systems-security and/or privacy*. 91–100.

[9] Digital Twin Consortium. 2023. Glossary. https://www.digitaltwinconsortium.org/glossary/glossary/#digital-twin

[10] CounterCraft. 2023. ActivebehaviorTM, Revolutionizing deception credibility. https://www.countercraftsec.com/blog/activebehavior-revolutionizing-deception-credibility/

[11] Violeta Damjanovic-Behrendt, Michaela Mühlberger, Cristina de Luca, Thomos Christos, Christoph Schmittner AIT, and Edin Arnautovic. 2018. Deliverable D5. (2018).

[12] DarkStax. 2023. Role of Digital Twin for Cyber Resiliency. https://darkstax.com/cyber-resiliency/

[13] Marietheres Dietz, Manfred Vielberth, and Günther Pernul. 2020. Integrating digital twin security simulations in the security operations center. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*. 1–9.

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

[15] CSRC Content Editor. [n. d.]. Honeypot - glossary: CSRC. https://csrc.nist.gov/glossary/term/honeypot

[16] Rajiv Faleiro, Lei Pan, Shiva Raj Pokhrel, and Robin Doss. 2022. Digital twin for cybersecurity: Towards enhancing cyber resilience. In *Broadband Communications, Networks, and Systems: 12th EAI International Conference, BROADNETS 2021, Virtual Event, October 28–29, 2021, Proceedings 12*. Springer, 57–76.

[17] Daniel Fraunholz, Daniel Reti, Simon Duque Anton, and Hans Dieter Schotten. 2018. Cloxy: A context-aware deception-as-a-service reverse proxy for web services. In *Proceedings of the 5th ACM workshop on moving target defense*. 40–47.

[18] Peter J Hanson, Lucas Truax, and David D Saranchak. 2018. IOT honeynet for military deception and indications and warnings. In *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*, Vol. 10643. SPIE, 296–306.

[19] David Holmes, Maria Papathanasaki, Leandros Maglaras, Mohamed Amine Ferrag, Surya Nepal, and Helge Janicke. 2021. Digital Twins and Cyber Security–solution or challenge?. In *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*. IEEE, 1–8.

[20] Thorsten Holz. 2005. Anti-honeypot technology.

[21] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 51, 11 (2018), 1016–1022.

[22] Efrén López-Morales, Carlos Rubio-Medrano, Adam Doupé, Yan Shoshitaishvili, Ruoyu Wang, Tifanny Bao, and Gail-Joon Ahn. 2020. HoneyPLC: A Next-Generation Honeypot for Industrial Control Systems. In *In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20)*.

[23] Tom H Luan, Ruhan Liu, Longxiang Gao, Rui Li, and Haibo Zhou. 2021. The paradigm of digital twin communications. *arXiv preprint arXiv:2105.07182* (2021).

[24] Forrest McKee and David Noever. 2023. Chatbots in a honeypot world. *arXiv preprint arXiv:2301.03771* (2023).

[25] Marco MELLIA and Danilo GIORDANO. 2021. Next Generation Honeypot for IoT. (2021).

[26] Maryam Mohammadzad and Jaber Karimpour. 2023. Using rootkits hiding techniques to conceal honeypot functionality. *Journal of Network and Computer Applications* 214 (2023), 103606.

[27] Valentin Mullet, Patrick Sondi, and Eric Ramat. 2021. A review of cybersecurity guidelines for manufacturing factories in industry 4.0. *IEEE Access* 9 (2021), 23235–23263.

[28] Maria Nintsiou, Elisavet Grigoriou, Paris Alexandros Karypidis, Theocharis Saoulidis, Eleftherios Fountoukidis, and Panagiotis Sarigiannidis. 2023. Threat intelligence using Digital Twin honeypots in Cybersecurity. In *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*. IEEE, 530–537.

[29] NIST. 2023. The NIST Cybersecurity Framework 2.0.

[30] Roland Rosen, Georg Von Wichert, George Lo, and Kurt D Bettenhausen. 2015. About the importance of autonomy and digital twins for the future of manufacturing. *Ifac-papersonline* 48, 3 (2015), 567–572.

[31] C. Sanders. 2020. *Intrusion Detection Honeypots: Detection Through Deception*. Applied Network Defense. https://books.google.no/books?id=suubzQEACAAJ

[32] Lance Spitzner. 2003. *Honeypots: tracking hackers*. Vol. 1. Addison-Wesley Reading.

[33] Shreyas Srinivasa, Jens Myrup Pedersen, and Emmanouil Vasilomanolakis. 2020. Towards systematic honeytoken fingerprinting. In *13th International Conference on Security of Information and Networks*. 1–5.

[34] Andrew D Syrmakesis, Cristina Alcaraz, and Nikos D Hatziargyriou. 2022. Classifying resilience approaches for protecting smart grids against cyber threats. *International Journal of Information Security* 21, 5 (2022), 1189–1210.
[35] Joni Uitto, Sampsa Rauti, Samuel Laurén, and Ville Leppänen. 2017. A survey on anti-honeypot and anti-introspection methods. In *Recent Advances in Information Systems and Technologies: Volume 2 5*. Springer, 125–134.
[36] EV Zavadskii and DV Ivanov. 2022. Counteracting Information Threats Using Honeypot Systems Based on a Graph of Potential Attacks. *Automatic Control and Computer Sciences* 56, 8 (2022), 964–969.
[37] Yuqiang Zhang, Zhiqiang Hao, Ning Hu, Jiawei Luo, and Chonghua Wang. 2022. A virtualization-based security architecture for industrial control systems. In *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*. IEEE, 94–101.
[38] Tianming Zheng, Ming Liu, Deepak Puthal, Ping Yi, Yue Wu, and Xiangjian He. 2022. Smart Grid: Cyber Attacks, Critical Defense Approaches, and Digital Twin. *arXiv preprint arXiv:2205.11783* (2022).
[39] Cheng Zhou, Hongwei Yang, Xiaodong Duan, Diego Lopez, Antonio Pastor, Qin Wu, Mohamed Boucadair, and Christian Jacquenet. 2021. Digital twin network: Concepts and reference architecture. *Internet Engineering Task Force: Fremont, CA, USA* (2021).

A A POC LLM-BASED HONEYPOT

This section describes DecEPt (**De**v**i**c**e** **E**mulation through **P**rompt**s**), a Proof of Concept (PoC) of a low-interaction honeypot based on a Large Language Model (LLM) inspired by the work of Mckee [24]. The PoC showcases the intersection between digital twins and cyber deception. We used the PoC to demonstrate an application scenario of the deception framework discussed in Section 4.

The honeypot operates as a low-fidelity digital twin of any OT device that uses a client-server architecture to communicate commands and responses over the network. The LLM drives the contextual information that allows the twin to respond accurately to appropriate commands. In this PoC, we used the ChatGPT 3.5 model. The honeypot is configured to log connections and notify defenders when a connection is made, demonstrating the deception aspect.

The high-level architecture of DecEPt is shown in Figure 4. DecEPt establishes a network presence as a Modbus server to attract attackers. When an engagement with an attacker occurs, DecEPt uses ChatGPT 3.5 to retrieve realistic yet fake responses as a typical device would respond. Additionally, DecEPt notifies the defenders of an ongoing engagement and logs the transactions. The prompt engineering element of DecEPt focuses on three aspects: context definition, command transfer, and response control. The context definition is achieved by defining the device that DecEPt should emulate. The command transfer step considers any information or restrictions to control the engagement scope with the attacker to be representative of the operational scope of the twinned device, for instance, the list of accepted commands by the device. Finally, the response control is achieved by restricting the models' response to a usable output that fits the communication channel.

The following is an explanation of how DecEPt realizes the core functions in the proposed deception framework:

- **Mimic**: DecEPt uses Transport Control Protocol (TCP) to create a communication channel that mimics devices, allowing it to send and receive commands as a real device would. To reflect a Modbus network device, port 502 is used. The device functionality is emulated by relying on the LLM capacity to provide relevant responses to commands.
- **Attract**: DecEPt advertises open ports as Modbus TCP service as decoys to attract attackers to engage.

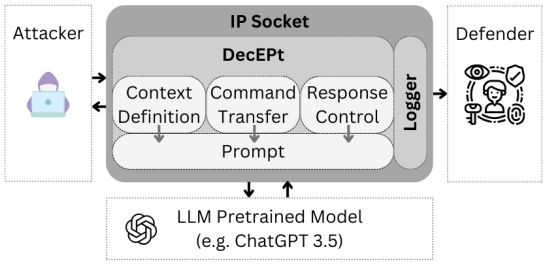


Figure 4: DecEPt Architecture

- **Deploy**: DecEPt can be deployed as a container or standalone script. It can easily resemble any device, which allows for mass deployment of a large network of OT devices.
- **Engage**: DecEPt establishes an engaging communication channel for attackers developed as a chat behind their knowledge. The engagement depends on the fidelity level the LLM model can provide based on its training.
- **Monitor**: DecEPt logs all transactions and notifies the defenders of any connections made.
- **Reset**: The reset function is implicitly implemented in the programming of the interface with the ChatGPT API. Each connection occurs in a fresh context, meaning that the status of the twinned device is always new.
- **Analyze**: DecEPt supports the analysis function by logging the attacker's engagements. All IP addresses, ports, commands, and responses are logged. This can help understand adversarial behaviour in the network better.

A.1 Demonstration of the PoC

We configured DecEPt to mimic MGate 5114, an Industrial Ethernet gateway. This can easily be changed and depends on ChatGPT training data. An engagement example is shown in Figure 5. In this example, an attacker establishes a TCP connection with the device, sends a "get device_info" command and receives a realistic response. Later, the attacker sends another command to read the coils' values and receives the response.

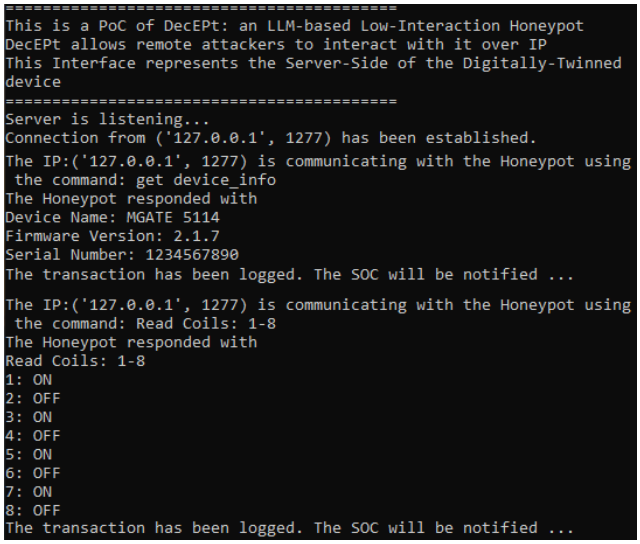


Figure 5: Screenshot of DecEPt interface reporting an engagement

Knowing the general nature of this model, practical limitations of DecEPt are to be expected and are observed in this work. Among the observed limitations of this implementation approach are the variation of the response formats by the model, the inability to identify irrelevant commands effectively, response delay, and the weakness of contextual information about the entire operational environment. Such restrictions might also impact the framework functions implementation, therefore, fine-tuning models for each device is needed and might achieve higher fidelity. Additionally, in a real operational network, offline models should replace online models. Lastly, time analysis should be considered during the development of such honeypots to avoid detection by attackers.