Journal of Innovation
May 2025 | 27th Edition

Threat Modeling for Digital Twins

Based on the DTC Platform Stack Architectural Framework

Author:

Ekaterina Rudina
Kaspersky
ekaterina.rudina@kaspersky.com

CONTENTS

1 Introduction 3
1.1 Digital Twin Trustworthiness and Security 3
1.2 Threat Modeling Approach 4
1.3 Threat Modeling Method 5
1.4 Case Study Under Consideration 8
2 Assets and Damage Considerations for the Attack on Digital Twins 9
3 Damage Categories and Ratings 13
4 Attack Vectors, Scenarios, and Attacks Likelihood Assessment 15
5 Considerations on Risks 16
6 Final Considerations 18
7 References 19
8 Acknowledgements 19

FIGURES

Figure 1-1: Threat modeling and risk assessment method in a nutshell. 7
Figure 1-2: The architecture of the FleetTwin case study. 9
Figure 5-1: Table with risk ratings. 17
Figure 5-2: Example of risk ratings. 17
Figure 5-3: Considerations on risks. 18

TABLES

Table 1-1: Threat modeling goals depending on the technology readiness levels of digital twins. 6
Table 2-1: The list of assets for the FleetTwin case study. 12
Table 3-1: Damage assessment for the FleetTwin case study. 14

1 INTRODUCTION

The digital twin is characterized by its ability to simulate and synchronize data between the virtual representation or acting model and the real world. This ability may become a specific threat vector. The complete system integration which makes us consider the acting model, application of platform as digital twin system may have an unrecognized impact on the real-world systems, especially in the case of cybersecurity attack. We'll consider the approach on threat analysis and assessment of risks to improve the assurance on proper digital twin implementation and use.

The proposed approach to threat analysis and cybersecurity risk assessment is aimed to improve trustworthiness of digital twins which is the benefit for business. The analysis is based on the known methods of risk assessment but also considers the specific purpose and capabilities of the digital twin system.

1.1 DIGITAL TWIN TRUSTWORTHINESS AND SECURITY

A digital twin is an integrated data-driven virtual representation of real-world entities and processes, with synchronized interaction at a specified frequency and fidelity. In terms of *system engineering*[1] approach, the digital twin is an *engineered system*[2] which comprises a part of more complex system with *emergent properties*.[3] These properties may contribute to the trustworthiness of the whole system or its parts, or, conversely, may pose threats.

Trustworthiness in the context of the Internet of Things vocabulary is defined as the "ability to meet stakeholders' expectations in a verifiable way" [1]. Depending on the context or sector, and on the specific product or service, data, technology and process used, different characteristics apply and need verification to ensure that stakeholders' expectations are met. Characteristics of trustworthiness include, for instance, accountability, accuracy, authenticity, availability, controllability, integrity, privacy, quality, reliability, resilience, robustness, safety, security, transparency and usability. Further use of the term "trustworthiness" in this article means "trustworthiness of digital twins" since this is the subject of our discussion. Trustworthiness of digital twins assumes the proper implementation of the intended digital twin purpose and the absence of harm to the environment, both digital and physical, even under cyberattack.

Security in the context of trustworthiness is the assurance that a protective measure is effective relative to an actual or perceived cyber threat. Among the consequences of threat, the impact on safety may be considered, and maintaining safety may be one of security objectives. Generally, security objectives usually reflect the system purpose and scenarios of its intended use. The Internet of Things document [2] mentions maintaining privacy while working online,

[1] *https://sebokwiki.org/wiki/Systems_Engineering_Overview*
[2] *https://sebokwiki.org/wiki/Engineered_System_(glossary)*
[3] *https://sebokwiki.org/wiki/Emergent_Property_(glossary)*

component security tolerance to reverse engineering, authenticity of the system software updates and other specific examples of security objectives. This puts the traditional "confidentiality, integrity, availability" triad far behind the current understanding of security in the context of trust.

This brings us to the following ideas:

- Importance of threat assessment, including threat modeling, for supporting trustworthiness of complex systems incorporating digital twins.
- Necessity of threat definition for digital twins in a wider context, consideration of possible system availability loss, safety impact or even physical damage.

1.2 THREAT MODELING APPROACH

Threat modeling is a structured, repeatable process used to gain actionable insights into the security characteristics of a particular system. Threat modeling analyzes a system from an adversarial perspective, focusing on ways in which an attacker can exploit a system.

There is no industry standard approach to threat modeling, nor is there a one-size-fits-all solution for all use cases. This is especially true for the Internet of Things, cyber-physical systems and full-fledged digital twins. Threat modeling considerations for these systems should include hardware/physical threats, interdependencies, scalability, and iterative processes throughout the system lifecycle.

To approach the issue of threat definition and further threat assessment, two *viewpoints*[4] on digital twin systems are needed. The first viewpoint is connected to the digital twin functional capabilities, according to which we determine possible damage from cyberattacks. Each capability may be either broken or abused because of cyberattack. This leads to the loss of expected outcome or causes the unexpected behavior which may not have the best effect for the whole system. This viewpoint helps with the assessment of possible damage to the digital twin system. The second viewpoint focuses on the digital twin architecture, components and technologies to reveal threat scenarios and evaluate attack likelihood. When combined, both viewpoints can help assess and mitigate cybersecurity risks.

The Platform Stack Architectural Framework [3] (further referred to as "framework") is suggested to form the base for the mentioned viewpoints. This framework is intended to provide a robust foundation for building secure, interoperable, and scalable digital twin systems and underscores the necessity of trustworthiness, system integration, and domain-specific customization to unlock transformative value.

[4] *https://sebokwiki.org/wiki/Viewpoint_(glossary)*

The framework considers and systematizes the capabilities of the digital twin ecosystem. This systematization along with description of the elements and sub-systems of digital twin, may help to evaluate the possible damage of cyberattacks.

Also, the framework describes the architecture approach and highlights how this approach relates to other architecture approaches. This helps to assess the attack surface more comprehensively, from multi-view perspective.

1.3 Threat Modeling Method

A fairly good overview of threat modeling methods is presented in [5] and in order to focus on digital twins, it will not be repeated here. Typical threat assessment processes are as follows:

- Considering possible targets of threats/objects of further assessment
- Determining the consequences of the threat implementation
- Assuming about the sources of threats
- Assessing possible threat methods based on the attack surface
- Evaluating attacking techniques, vulnerabilities based on the system-related details, and
- Assessing possible attack scenarios by summarizing all details above.

A good threat modeling method focuses on an object of assessment with particular capabilities and well-defined attack surface. According to the level of maturity defined for digital twins defined in McKee's white paper [3] and the appropriate technology readiness level (TRL)[5], the range of capabilities may vary drastically. Similarly, as a digital twin becomes a digital twin system, its attack surface becomes wider and much more complex. The adversary may have significantly different goals when attacking digital twins of varying degrees of maturity.

Thus, we have to determine threat modeling goals and objectives depending on the different technology readiness levels, as they apply to digital twins (Table 1-1).

TRL	Threat Modeling Goals
1-3 (Research)	• Protect the data used for research and development of digital twin from leakage • Protect the data and algorithms used for research and development of digital twin from manipulation • Avoid inaccuracies and logical errors in models and algorithms, which may be used to abuse digital twin scenarios

[5] Technology Readiness Levels which help to understand technical maturity of a system are used here in the same way as they used in the framework. More about TRL scale: *https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/*

4-6 (Development)	• All goals defined for TRL 1-3 • Ensure proper synchronization enabling feedback and control between the digital twin and the physical entity, protect synchronization from unauthorized interference • Avoid architectural issues and vulnerabilities while integrating with IT/OT systems (e.g. IoT devices, enterprise platforms)
7-9 (Deployment)	• All goals defined for TRL 1-6 • Ensure resilience to cyberattacks for a full-fledged digital twin system: even if one of the components is compromised, the entire system maintains availability of services and remains trustworthy

Table 1-1: Threat modeling goals depending on the technology readiness levels of digital twins.

Threat modeling method and objectives may be further detailed based on the lifecycle stage, or process under which the assessment is conducted. For example, for the application under development, the STRIDE method[6] is applied to reveal the issues of the particular categories (spoofing, tampering, etc.) and prevent them regardless of the damage for the system under attack.

During development, the essential impact of tampering with data or denial of service of a separate piece of code may not be clear for the entire system when it is subsequently used. This is just a good practice to prevent the code issues to anticipate and prevent any effect they may have in future. For the system under design, the attack trees[7] method may be used to analyze threats from the general to the specific ones. This approach also enables the connection of threats to attacking techniques and vulnerabilities at the latest stage of design.

While different threat modeling techniques can be used to assess threats and analyze risks in different SDLC processes of a digital twin (and its components), we will focus on a threat model that will support trustworthiness to the digital twin in terms of the emergent values it brings.

These values are described in McKee's white paper [3] as follows:

- Business transformation by accelerating holistic understanding, optimal decision-making and effective action,
- Ability to represent the past and present and simulate predicted futures based on both real-time and historical data,

[6] STRIDE abbreviation stands for spoofing, tampering, repudiation of origin, information disclosure, denial of service, elevation of privilege. STRIDE helps to identify cybersecurity threats and used for threat modeling in conjunction with the description of the system under attack.

[7] Attack tree is the diagram showing how the target might be attacked. The root of the tree describes the general threat and the leaves detail the conditions under which the attack is implemented.

- Fine adjustment based on multiple factors: motivation by outcomes, tailoring to use cases, integration, data, domain knowledge and particular implementation in IT/OT systems.

Security risks for the digital twin system are connected to the violation of capabilities determining system behavior and supporting these values. Each capability is implemented by one or more components of composable digital twin. Therefore, considering the attack scenarios on the components, we can assess both possible damage of an attack and its likelihood based on scenario.

Thus, threat modeling for a digital twin can be performed in two stages: assessing the potential damage through assessing the capability violation, identifying components related to capabilities, and assessing attack scenarios that can lead to the corresponding disruption of the component and the implemented capability. Finally, through the introduced qualitative or quantitative assessment of the damage and the probability of the attack scenario, it is possible to evaluate the level of risk. The entire process is illustrated in Figure 1-1.

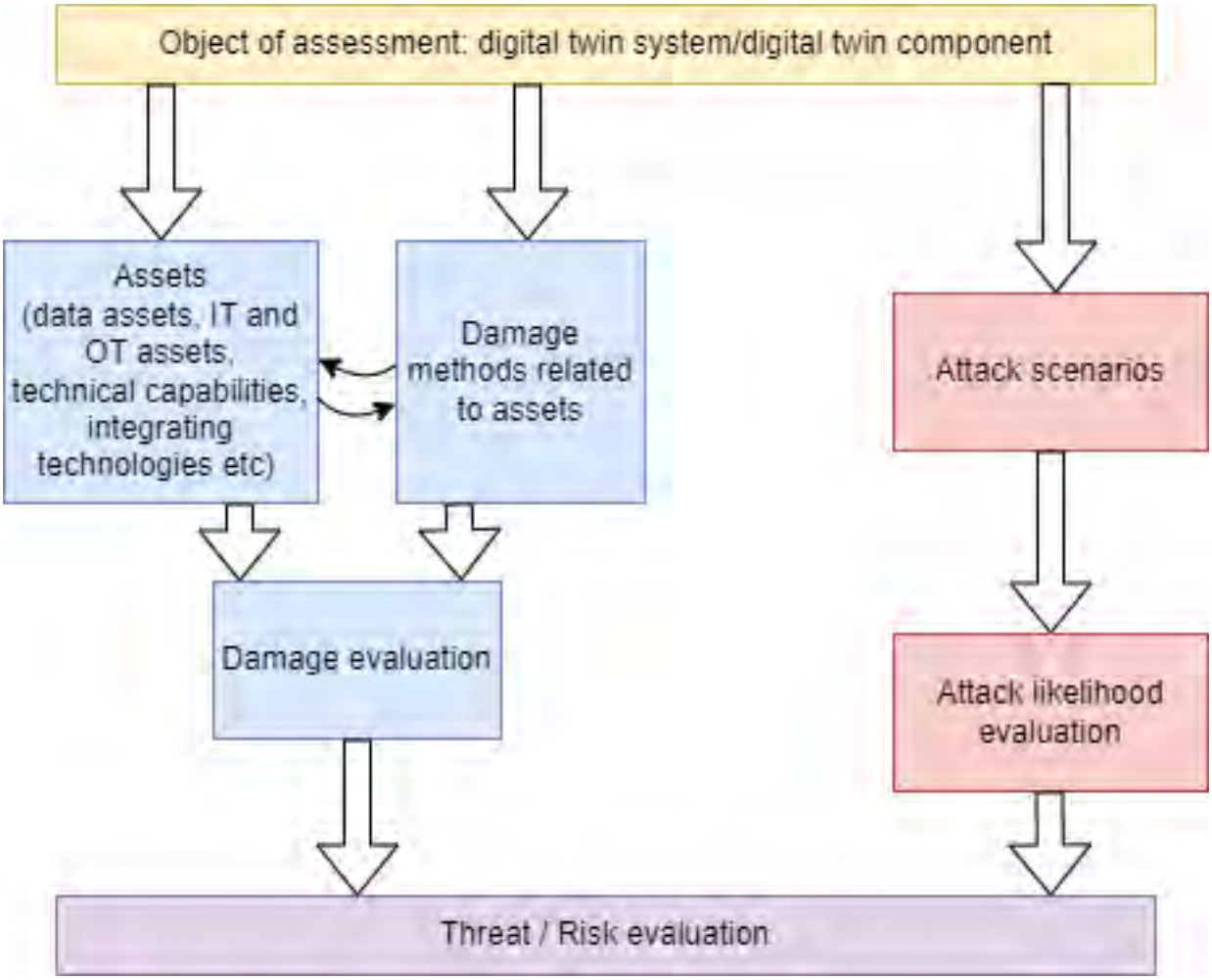

Figure 1-1: Threat modeling and risk assessment method in a nutshell.

1.4 CASE STUDY UNDER CONSIDERATION

To illustrate the proposed method, we will consider the digital twin using vehicle telematics data for the commercial fleet maintenance (hereinafter FleetTwin for short). Its primary goals are to:

- Monitor real-time vehicle geolocation and health (e.g., engine performance, battery status, tire pressure).
- Predict mechanical failures using predictive analytics to minimize downtime.
- Optimize maintenance schedules by analyzing historical and real-time data.
- Reduce operational costs through proactive repairs and fuel efficiency insights.
- Enhance decision-making for fleet managers via actionable insights (e.g. provide the nearest car service location in case of possible breakdown, suggest that the driver needs to be replaced etc.).

The following components comprise FleetTwin system:

1. In-vehicle telematics devices. These are electronic control units (ECUs) embedded in vehicles to collect data (e.g. GPS location, engine diagnostics, fuel consumption, brake wear) and transmit data to the cloud via cellular/satellite networks.
2. Data Transmission Network: cellular network connection via private APN (Access Point Name) to provide dedicated network access, reduce exposure to public networks, and enable control over data traffic. Ensures connectivity between vehicles and the cloud platform.
3. Cloud Storage & Processing Platform, which stores raw telematics data (e.g. AWS, Azure, Outscale) and preprocesses data (cleaning, normalization) for analysis.
4. Digital Twin Engine, which generates and updates virtual vehicle models using real-time data and simulates scenarios (e.g. stress testing components, failure modes).
5. Analytics & Machine Learning (ML) Module to apply ML algorithms to detect anomalies and predict failures (e.g. engine breakdowns), integrate historical data, OEM specifications, and environmental factors (e.g. weather).
6. User Interface (Dashboard) based on web interface/mobile application for fleet managers to report vehicle location data, view vehicle health, alerts, and provide recommendations. Enables manual overrides (e.g. rerouting vehicles to the car service for urgent repairs).
7. Maintenance Management Integration component: automatically schedules repairs via integrations with enterprise systems (e.g. SAP, Oracle) and shares work orders with repair centers and parts suppliers.
8. Third-Party Services enabling external APIs for weather data, traffic updates, or OEM diagnostic tools.

For this case study, we don’t consider feedback of digital twin system to the real world and possible safety impact. It is also assumed that all data from vehicles don’t contain any private data. Generally, security issues are in focus for trustworthiness (see Figure 1-2).

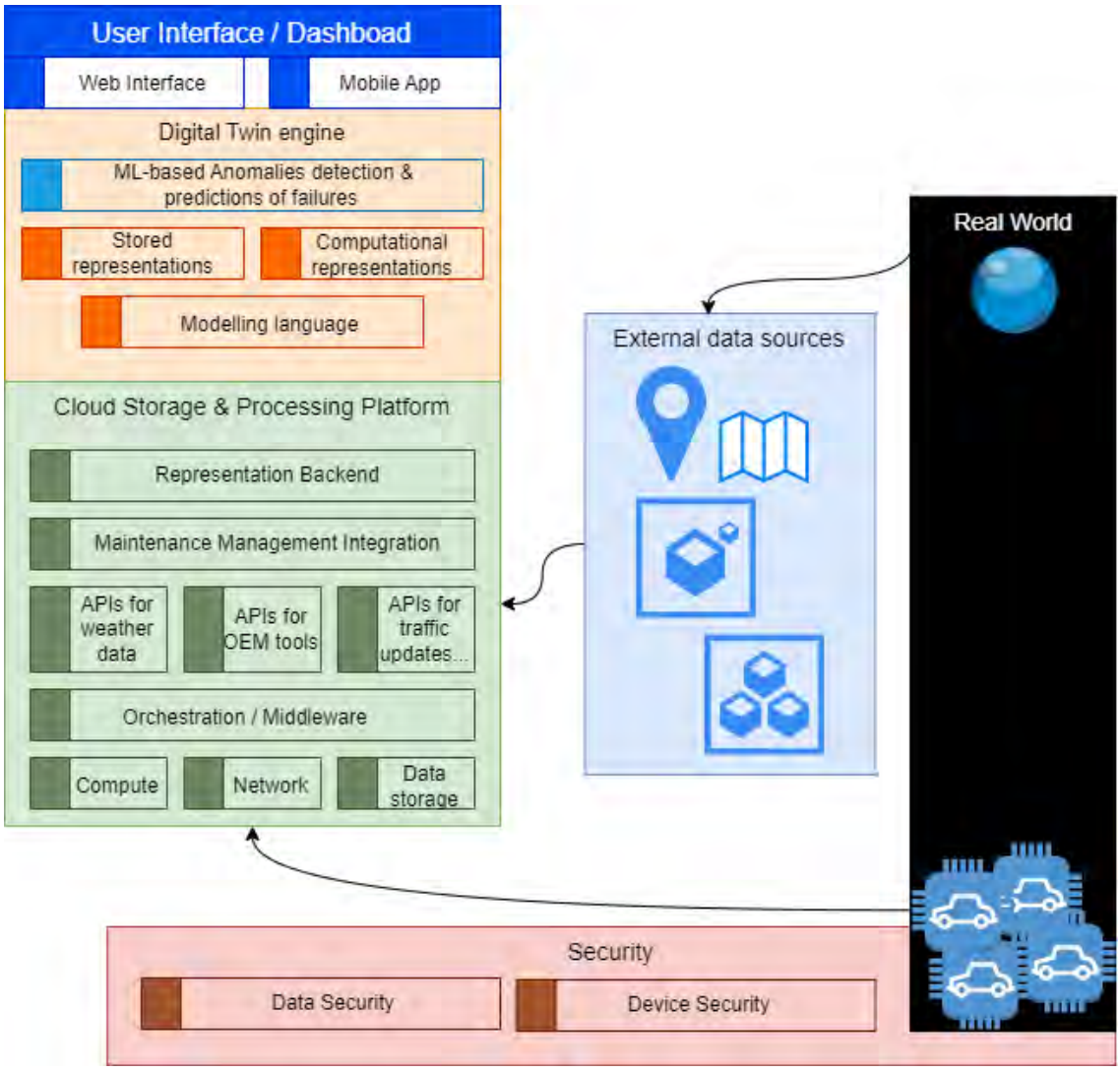

Figure 1-2: The architecture of the FleetTwin case study.

2 Assets and Damage Considerations for the Attack on Digital Twins

Listing the assets is essential for many methods of threat modeling. At the same time, it may be challenging when it comes to applying it to a real case. Asset identification is the minor issue if you need just to label them for registration. Assets in threat modeling are connected to damage scenarios, then – to the attacks and attacking scenarios, and then – to vulnerabilities. Each step multiplies the number of entities under consideration and makes the threat model too vast and complex. That means we need to reduce these numbers at each step, starting from the number of assets.

Main issues of asset identification for threat modeling are the following:

- Granularity (e.g. what is the asset: the data storage or the separate file)

- Proper identification of assets when it comes to availability/security/other systems -ility to not related to a single component, but in the relationship of components
- Consistency of the assets.

The capabilities list from van Schalkwyk's user guide [4] may provide the base for asset identification, addressing the listed issues. The capabilities of digital twins, can be broken into six key high-level categories:

- Data services
- Integration
- Intelligence
- User experience (UX)
- Management and
- Trustworthiness.

Each of these categories further expands into 62 discrete top-level capabilities of a digital twin ecosystem.

The proposed approach for asset identification is to verbally describe the system and decompose it into the presented capabilities from the list. This helps not to duplicate assets and not to miss anything in their listing.

Example

By using this approach, the following capabilities from [4] are listed for the FleetTwin case study:

1. Specific devices (ECUs) are used for data gathering from each vehicle, these devices and connected sensors are critical for the entire system so they are under monitoring:
    - Device Security
    - Device Management
    - Event logging
    - System monitoring
2. Telematics devices continuously send data to the cloud via secured network connection:
    - Data Acquisition & Ingestion
    - Data Streaming
    - Data Encryption
3. Raw data is cleaned, structured, and stored in databases:
    - Data Transformation
    - Data Contextualization
    - Batch Processing
    - Data Aggregation
4. Digital twin engine pulls the data to update virtual models in near real-time.

    - Real-time processing
5. ML models analyze processed data to flag anomalies (e.g. overheating engine):
    - Machine Learning
    - Digital Twin (DT) Model Repository
    - Data Analysis & Analytics
6. Fleet managers have access to dashboards and detailed information on separate vehicles using web interface / mobile applications:
    - Basic Visualization
    - Dashboards
    - Reporting
7. Fleet managers review alerts and approve automated maintenance requests:
    - Prescriptive Recommendations
8. Predictions trigger alerts in the user dashboard and maintenance systems.
    - Prediction
    - Alerts & Notifications
9. Maintenance schedules are pushed to third-party repair systems via APIs.
    - Enterprise System Integration
    - Collab Platform Integration
    - API Services
10. Post-repair data from service centers updates the digital twin, refining future predictions:
    - Digital Twin Integration

The analysis starts from the consideration of each capability as the asset. Then, if the capability is connected to other kinds of assets, more convenient for analysis, it may be replaced with that asset. For example, privacy (which is a capability considered as an asset) can be replaced with PII (as a data asset). Data-related capabilities can be replaced with the data they control or process. In some cases, an abstractly described capability (such as machine learning) can be replaced by a set of functions with a specific purpose (e.g. ML based anomaly detection for enhanced network security).

Various capabilities, representing the generic valuable functionality (like maintenance functionality), may be grouped together to form the asset. The main criteria for proper asset identification are the ability to connect the asset to possible damage in case of cyberattack or other kind of failure or disruption, to name and assess this damage. The empirical analysis of the damage kind (what can happen to the asset) is enough to validate the asset presence in the list.

Example

By following these recommendations, we get this list of assets with connected damage kinds (Table 2-1).

Asset Description	Capabilities transformed to the asset	What can happen to the asset
Telematics device (ECU) on vehicle	Device Security Device Management	Physical tampering, removal Firmware/software tampering or invalid update
Raw data on vehicle	Data Acquisition & Ingestion Data Streaming Data Encryption	Raw data tampering
Processed data in cloud	Data Transformation Data Contextualization Batch Processing Data Aggregation	Abuse of processing algorithms Processed data tampering Processed data removal
Service data on vehicle and in cloud	Event logging System monitoring	Service data tampering Service data removal
Seamless integration capability to reduce service costs	Real-time processing Prediction Digital Twin Integration Enterprise System Integration Collab Platform Integration API Services	Abuse of real-time processing and prediction algorithms Denial (delay) of service of external platforms and services Spoofing of external platforms and services Tampering of data from external services
ML capability to flag anomalies	Machine Learning Digital Twin (DT) Model Repository Data Analysis & Analytics	Abuse of machine learning algorithms (adversarial inputs, data poisoning) Model stealing
Maintenance capability	Basic Visualization Dashboards Reporting Prescriptive Recommendations	Denial of service of visualization and maintenance services Data tampering for visualization and reporting Spoofing of recommendation services

Table 2-1: The list of assets for the FleetTwin case study.

The predefined list of capabilities of digital twin helps to define a pretty short list of valuable assets. This list may be validated with stakeholders and used to describe damage from attacks. At the same time, the clear connection to capabilities can be used for further analysis to reveal attack paths.

3 DAMAGE CATEGORIES AND RATINGS

The damage is directly related to risk. Measuring the possible impact of a cyberattack makes it possible to assess the risk in financial or other terms. For Industrial Internet of Things (IIoT) systems, including digital twins, the possible consequences of attacks are not only related to financial losses, but also to the operational consequences, sometimes to safety and other trustworthiness aspects.

Damage categories help to determine risk measurement approach. Damage is also directly connected to the assets which we just defined. By reviewing the short list of assets, stakeholders can define damage categories and ratings by following the recommendations:

- Financial category is preferable: if the damage can be measured by money or equivalent, the appropriate method should be used.
- Categories of damage used by the industry: incidents that can happen at least to physical assets are usually measured by their severity; these scales can be appropriately reused for the cyberattacks impact assessment.
- Newly introduced categories and ratings of damage or the neutral rating like "high/medium/low damage" may be used only if existing ones do not fit the nature of possible threat.

Example

The FleetTwin case study can be analyzed for potential adverse consequences using ISO/SAE 21434:2022, which includes impact categories such as safety, financial, operational, and privacy (S, F, O, P). Damage ratings are introduced as follows:

1. Safety impact rating criteria are taken from *ISO 26262-3:2018*:
    a. Severe (S3) - Life-threatening injuries (survival uncertain), fatal injuries
    b. Major (S2) - Severe and life-threatening injuries (survival probable)
    c. Moderate (S1) - Light and moderate injuries
    d. Negligible (S0) - No injuries
2. Financial impact rating criteria are connected to the number of cars in the fleet and the average annual cost of service.
    a. Severe (S3) – Exceeding the estimated fleet maintenance costs by more than 100% over a certain period of time (at least one month)
    b. Major (S2) - Exceeding the estimated maintenance costs by more than 50%
    c. Moderate (S1) - Exceeding the estimated maintenance costs by more than 10%
    d. Negligible (S0) - Exceeding the estimated maintenance costs by less than 10%
3. Operational impact rating criteria are connected to the number of cars in the fleet and the time required for the car to be repaired due to the improper maintenance (comparing to the optimized predicted time).

   a. Severe (S3) – Total fleet maintenance time exceeded by more than 100% over a certain period of time (at least one month)
   b. Major (S2) - Total fleet maintenance time exceeded by more than 50%
   c. Moderate (S1) - Total fleet maintenance time exceeded by more than 10%
   d. Negligible (S0) - Total fleet maintenance time exceeded by less than 10%
4. Privacy impact rating criteria are not considered applicable following the analysis, since the case study does not involve the processing of private data and does not interact with PII holders and other systems that process PII.

Percentage for the financial and operational impact is indicative and subject to change based on stakeholders' opinions. The other type of risk - business risks - comes into play here.

Damage methods are considered then for the assets. Damage method is connected to the threat scenario. It's important to distinguish between the damage method and the attack scenario. Damage method is the action that becomes possible because of the attack, the adverse consequence of it. Each damage method for the asset can be evaluated using the damage categories and ratings: the worst consequences of threat for the object under analysis. Example of damage methods assessment for the FleetTwin case study is provided below.

Example

Asset	Damage method	Safety impact	Financial impact	Operational impact
Telematics device (ECU) on vehicle	Firmware/software tampering	Moderate	Major	Severe
	…			
Seamless integration capability to reduce service costs	Denial (delay) of service of external platforms and services	Negligible	Moderate	Moderate

Table 3-1: Damage assessment for the FleetTwin case study.

Impact evaluation is justified as follows.

Example

Firmware tampering for the telematics device may have safety consequences. Insufficiently validated software may affect the operation of other ECUs in the vehicle Electrical/Electronic (E/E) network. At the same time, we assume that the safety requirements for the E/E architecture address overall reliability and safety, even in case of failure of some of the devices in the network. We also assume that improper device may cause separate failures and glitches of the vehicle equipment and lead to the light and moderate injuries in case of car accident (moderate safety impact). Financial impact in

> major, and operational impact is severe because of the need of ECUs replacement or firmware update at the service station.
>
> Seamless integration to reduce service costs is one of the main goals of FleetTwin solution. If one of external services supporting this capability fails, it may cause moderate financial and operational impact, but will not affect safety.

This evaluation becomes more formal as the details come. The worst conditions should be evaluated, but coincidences and combinations of events are usually not considered.

4 Attack Vectors, Scenarios, and Attacks Likelihood Assessment

When damage methods for the assets are evaluated, the likelihood of attacks should be assessed. The typical approach is to consider the attack surface, possible attack paths, describe attack scenarios and the likelihood of each scenario. Although the preceding stage does not require specialized cybersecurity knowledge, it is best to have cybersecurity experts perform these steps.

This is the other viewpoint, focusing not on the system capabilities but on the technical interdependencies between assets, attackers, methods, tools, and attack surfaces.

The best way to define the attack surface is to consider system architecture and external interfaces. Digital twin is the system-of-systems, comprised by at least two components.

Example

> Architecture of the FleetTwin system introduces multiple attack surfaces:
> - In-vehicle Telematics Devices: Physical tampering or spoofing sensor data.
> - Data Transmission: Man-in-the-middle attacks on unsecured networks.
> - Cloud APIs: Exploitable vulnerabilities enabling data tampering and denial of service.
> - Third-Party Integrations: Compromised suppliers could inject malicious data.
> - User Access: Weak authentication might allow unauthorized dashboard access.

Attack scenario is the set of deliberate actions to implement the threat. Possible attack scenarios are identified at the intersection of the description of system interfaces, system architecture and assumptions about the attacker. These assumptions include, for example, physical location, capabilities, knowledge and motivation.

Assessing attack scenarios for digital twin infrastructure is essentially the same as assessing other types of systems and networks.

The first step is to identify the attack vector: how the attacker accesses the system or its component. For the complex systems, depending on the access vector, the attacker may plan further actions and follow a tactic to eventually implement the damage method. This may be evaluated based on *MITRE Att&ck*[8] matrices (or similar matrices of attacking tactics and techniques). Tactical schemes vary depending on the system purpose and typical internal

[8] *https://attack.mitre.org/*

organization (e.g. ICS matrix is separate from the matrix for the enterprise environment). The attacking techniques are implemented at each step, work because of the technology issues and vulnerabilities of components. Actually, the attack likelihood may be evaluated depending on the complexity of the supposed tactic, each attacking technique and exploited vulnerabilities.

The disadvantage of this method is its complexity and the need to know the implementation details of the system. When it is unclear whether the technique may be used to attack the component, we have to follow the pessimistic approach (all techniques are applicable) or make the assumptions which are probably inappropriate. At the same time, this is the most optimal way to describe and evaluate attack scenarios for the complex environments.

Other approaches to assess the attack likelihood may be applied, for example:

- Based on the attack vector. This approach is used at the early stages of system design (when there are no details about system implementation). The likelihood of the remote attack is higher than the likelihood of the attack restricted by the physical access to the system.
- Based on the CVSS rating[9] of vulnerabilities used for the attack. This approach may fit the needs to assess likelihood of attack to the separate system components or platforms with known security issues.

Example

To describe attacks and assess attack likelihood for the FleetTwin case study, it is needed to consider both cloud environments and physical devices which may be exploited or physically tampered. It is difficult to decide whether we should use a single method and lose the benefits of knowing specific vulnerabilities or attack vectors, or use different ways of assessing attacks and end up with an inconsistent likelihood score for different components.

Finally, it is proposed to assess attacks on control units in the physical world based on knowledge of attack vectors and typical vulnerabilities of ECUs and the E/E architecture of the vehicle. For the cloud components, including machine learning, it is necessary to apply an assessment based on a matrix of attack tactics and techniques. For the simplicity, we unify the scale and refer the “low/medium/high” likelihood of the attack. For the external services accessed through APIs, we always assume the highest level of attack likelihood.

5 CONSIDERATIONS ON RISKS

After assessing the categories and size of potential damage and the likelihood of an attack, a risk rating must be assessed.

[9] The Common Vulnerability Scoring System (CVSS) is a method used to supply a qualitative measure of vulnerability severity rating.
https://nvd.nist.gov/vuln-metrics/cvss

At first glance, there will be no particular differences between digital twins and other types of IIoT systems. Usually, the rating is assessed based on a table that correlates the values of potential damage and the probability of an attack. It is quite easy to determine the maximum and minimum risk ratings in the corners of the table (see Figure 5-1).

		Damage rating			
		Severe	Major	Moderate	Negligible
Likelihood rating	Almost certainly				
	Very likely				
	Likely				
	Rather unlikely				
	Unlikely				

Figure 5-1: Table with risk ratings.

The question remains how to assess the risk rating in the cells in the middle of the table. Will a very likely risk with catastrophic consequences have the same degree as an “almost certain” risk with critical consequences? What degree of risk is acceptable, and will it be located in the middle of the table? What degree should be attributed to a very likely risk with minor consequences?

The point is that here in our table with a fairly simple scale of damage and a clear scale of five likelihood ratings there will be 20 risk level options (and they may also depend on the risk category). One of the examples is shown in Figure 5-2.

		Damage rating			
		Severe	Major	Moderate	Negligible
Likelihood rating	Almost certainly	Critical	Critical	High	Negligible
	Very likely	Critical	High	Moderate	Negligible
	Likely	High	High	Moderate	Negligible
	Rather unlikely	Moderate	Moderate	Low	Negligible
	Unlikely	Low	Low	Negligible	Negligible

Figure 5-2: Example of risk ratings.

Risk tolerance at this stage sometimes is determined empirically. No matter what symmetrical or asymmetrical scheme the analyst has determined, a refuting example may be found in life, because the scheme is only in the analyst's head. It needs reinforcement from real life.

Example

For example, firmware tampering for the telematics electronic control unit (ECU) on vehicle may have a *severe* operational impact. The attack on the firmware-over-the-air (FOTA) updating mechanism is assessed as *likely* because of the leaked cryptographic keys currently used to sign the updates. The risk is *high*, and the cybersecurity experts confirm this rating.

Denial of service of the manufacturer's cloud platform (accessed through an API) which supports the seamless integration capability for the digital twin have the *moderate* operational impact, and we have assumed, that it is *almost certainly* vulnerable to attack. The risk is also assessed as *high*. Not all experts will confirm it, but considerations on some service level agreement (SLA) provisions about service availability may be used to revise the rating.

General considerations for an analyst when conducting a risk assessment are summarized in Figure 5-3.

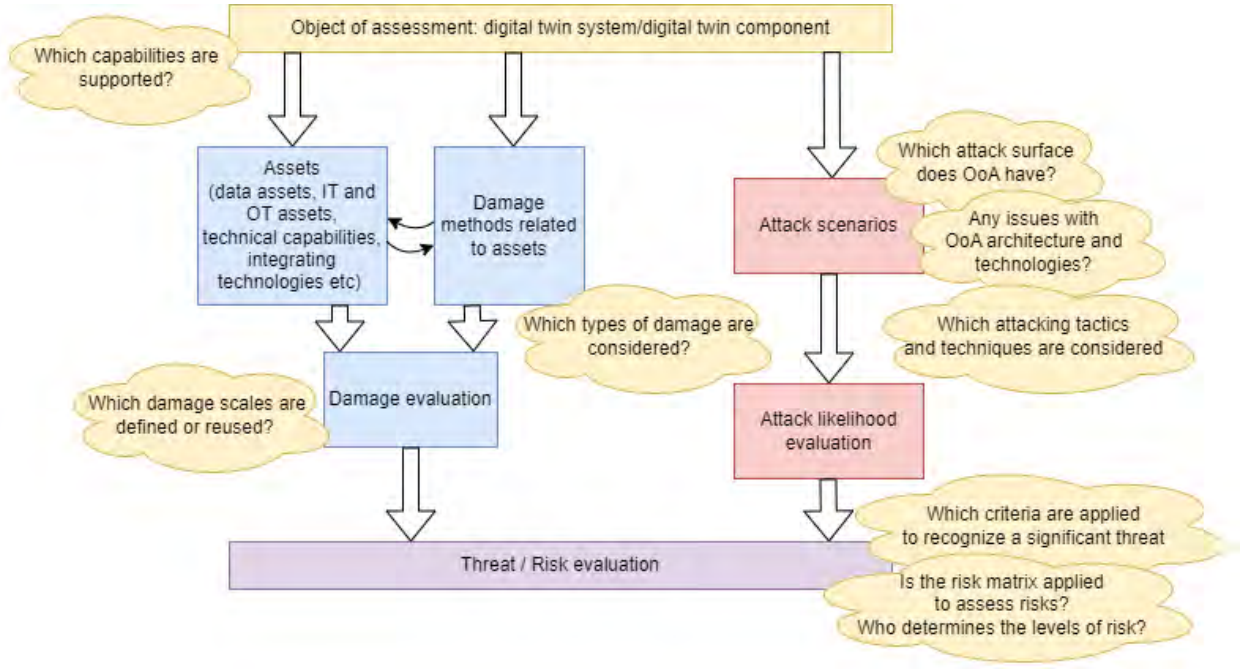

Figure 5-3: Considerations on risks.

6 FINAL CONSIDERATIONS

Capabilities supporting trustworthiness of the digital twin system should be in focus from the early stages of the concept phase. At the same time, not all details are available to assess hazards and threats and evaluate risks at early stages of the lifecycle. For the integrated system there are

too many details, and it may not be clear how to handle the information about the separate vulnerabilities and components exposed to attacks to get a proper risk rating.

The consistent approach based on the in-field knowledge and understanding of capabilities of the designed solution makes it possible to assess the risks. The threat model is updated together with the digital twin system for which it serves, maintaining the required level of security and trustworthiness assurance. This approach, therefore, contributes to security-by-design and security-by default principles of digital twin systems implementation, as described in [6].

7 REFERENCES

[1] ISO/IEC TS 5793:2022 Trustworthiness — Vocabulary

[2] ISO/IEC TS 30149:2024 Internet of Things (IoT) — Trustworthiness principles

[3] McKee, D. (2023). Platform Stack Architectural Framework: An Introductory Guide. A Digital Twin Consortium White Paper.

https://www.digitaltwinconsortium.org/wp-content/uploads/sites/3/2023/07/Platform-Stack-Architectural-Framework.pdf

[4] Pieter van Schalkwyk. (2022). Digital Twin Capabilities Periodic Table. A Digital Twin Consortium User Guide. *https://www.digitaltwinconsortium.org/wp-content/uploads/sites/3/2022/06/Digital-Twin-Capabilities-Periodic-Table-User-Guide.pdf*

[5] Shevchenko, N. (2019). Evaluating Threat-Modeling Methods for Cyber-Physical Systems

https://insights.sei.cmu.edu/blog/evaluating-threat-modeling-methods-for-cyber-physical-systems/

[6] Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-by Design and -Default. Cybersecurity and Infrastructure Security Agency (2023). *https://www.cisa.gov/sites/default/files/2023-04/principles_approaches_for_security-by-design-default_508_0.pdf*

8 ACKNOWLEDGEMENTS

The views expressed in the *OMG Journal of Innovation* are the author's views and do not necessarily represent the views of their respective employers nor those of the Object Management Group® (OMG®).

➢ Return to *OMG Journal of Innovation landing page* for more articles and past editions.