




**Digital
Water
.City**

Security assessment of cyber-physical flow of information in strategic, tactical and operational dimensions regarding DWC digital solutions



Deliverable N°	D4.3
Related Work Package	Work Package 4
Deliverable lead	Guillaume Bour
Author(s)	Rita Ugarelli, SINTEF Camillo Bosco, SINTEF Martin Gilje Jaatun, SINTEF Guillaume Bour, SINTEF
Contact for queries	Guillaume Bour (Guillaume.Bour@sintef.no)
Grant Agreement Number	n° 820954
Instrument	HORIZON 2020
Start date of the project	01 June 2019
Duration of the project	42 months
Website	www.digital-water.city
License	 <p>This work is licensed under a Creative Commons Attribution 4.0 International License</p>
Abstract	<p>This report builds on D4.1 and D4.2 to provide digital solution providers and water utilities with the tools and means to perform security assessments of the cyber-physical flow of information in strategic, tactical, and operational dimensions.</p> <p>We developed a risk guide aiming at supporting water organisations in the implementation of a complete risk management process in order to increase the preparedness of water operators against potential cyber and/or physical threats connected to the digital solutions adopted. In addition, a series of security recommendations and best practices for technology providers are described.</p> <p>Two digital solutions served as use cases to test and validate the risk guide and the security black box testing methodology presented as part of the security recommendations.</p> <p>Overall, this document describes a full process for secure integration of digital solutions within water utilities.</p>

Dissemination level of the document

<input checked="" type="checkbox"/>	PU	Public
<input type="checkbox"/>	PP	Restricted to other programme participants
<input type="checkbox"/>	RE	Restricted to a group specified by the consortium
<input type="checkbox"/>	CO	Confidential, only for members of the consortium

Versioning and contribution history

Version*	Date	Modified by	Modification reasons
D1	January 2022	Rita Ugarelli	Proposed Structure.
D2	March/April 2022	SINTEF	Contributions from the authors.
R1	09/05/2022	Guillaume Bour	Sent for internal review.
R2	27/05/2022	Nico Caradot	Review by coordinator.
S	30/05/2022	Guillaume Bour	Final version ready for submission.

* The version convention of the deliverables is described in the Project Management Handbook (D7.1). *D* for draft, *R* for draft following internal review, *S* for submitted to the EC (under external review) and *V* for approved by the EC.

Note that previous versions to *V* are draft since they are not yet approved by the EC.

Table of content

1.	Introduction	9
2.	The strategic and tactical decision level: how to perform a risk management process against cyber-physical threats	10
2.1.	Aim of the Risk Guide	10
2.2.	Content of the Risk Guide.....	11
2.3.	Target users of the Risk Guide	13
2.4.	Scope of use of the Risk Guide	13
3.	The operational decision level: general security recommendations to digital solutions.....	14
3.1.	Risks to digital solutions in DWC	14
3.2.	Threat landscape for the digital solutions	18
3.3.	Security testing methodology	19
3.3.1.	Penetration testing.....	19
3.3.2.	Red team exercise	21
3.4.	Recommendations to digital solutions’ owners	21
3.4.1.	General recommendations.....	21
3.4.2.	Attacks on sensors/IoT	26
3.4.3.	Attacks on servers/cloud	26
3.4.4.	Attacks on ML/AI	27
3.4.5.	Attacks on applications	27
3.4.6.	Attacks on humans and human errors	28
3.5.	Security assessment of a DWC digital solution	28
3.5.1.	Methodology	28
3.5.2.	Scope	30
3.5.3.	Results	31
3.5.4.	Exploitation.....	31
4.	Conclusion.....	31
	References.....	32
	Appendix A – Risk Management Guide for Cyber-Physical Attacks in Water Systems	35
A.1.	Introduction	35
A.2.	Defining the context	38
A.3.	Risk Identification	42
A.4.	Risk Analysis	44
A.5.	Risk Evaluation	56
A.6.	Risk Treatment.....	58
A.7.	ANNEX - InfraRisk-CP assessment for physical attacks (from the user guide of STOP-IT)	61
	Appendix B	64
	IoT Security Checklist	64

Server & Cloud Security Checklists & Requirements	70
Checklist for security in water process control networks	70
Cloud Security Requirements for Critical Infrastructure	72
ML & AI Security Checklist	77

List of figures

Figure 1. Risk management process as defined in ISO 31000:2018	9
Figure 2. Example of bow-tie diagram related to the relationship between a contingency and its causes and consequences.....	12
Figure 3 Generic Architecture Diagram for a Digital Solution in DWC	16
Figure 4 Classification of the attacks against DWC Digital Solutions.....	18
Figure 5 High level diagram of the Black Box Testing Methodology	20
Figure 6 ML System architecture (adapted from [28])	27
Figure 7 Black Box Methodology iterative cycles used during our assessment	30
Figure A.1. The Risk Management framework (ref: ISO 31000:2018).....	37
Figure A.2. Measured flows at the entrance of the analyzed WWTP in the year 2020	47
Figure A.3. Average daily flow pattern in dry weather conditions	48
Figure A.4. Estimated dilution coefficients at the entrance of the analyzed WWTP in the year 2020 ..	48
Figure A.5. Hourly flow values in 2020 greater than 7500 m ³ /h and coefficient dilution less than 3...	49
Figure A.6. Hourly values to be stored in 2020 with flow greater than 7500 m ³ /h and coefficient dilution less than 3	50
Figure A.7. Critical hourly values in 2020 not diluted, not stored, and not biologically treated.....	50
Figure A.8. Critical hourly values in 2020 of wastewater without any rain contribution.....	51
Figure A.9. Rain data in 2020 of seven station of the Danish rain gauge network.....	52
Figure A.10. Risk reduction measures in the RRMD associated with the identified risk in the RIDB. ...	59
Figure A.11. Critical hourly values in 2020 of wastewater without any rain contribution in the case of doubling the volume of the equalization tanks as risk reduction measure.	60
Figure B.1. Filled-in checklist with high-level overview	70

List of tables

Table A.1. Characterization of the features of the case study.....	37
Table A.2. An example of risk matrix with 5 levels of Risk.....	40
Table A.3. Characterization of the considered water system.....	40
Table A.4. Pre-defined level of Risk expressed in terms of the selected KPI.....	42
Table A.5. Identification of the level of risk by comparing results with targets values.....	57

Glossary

CI	Critical Infrastructure
CSO	Combined Sewer Overflow
DWC	Digital Water City
D	Deliverable
DS	Digital Solution
KPI	Key Performance Indicator
RIDB	Risk Identification Database
RRM	Risk Remediation Measures
RRMD	Risk Reduction Measures Database
WP	Work Package
WWTP	Waste Water Treatment Plant

Executive summary

Deliverable (D) 4.3 is the third of five deliverables, produced by Work Package (WP) 4 of DWC, related to identifying and mitigating risks arising from cyber-security viewpoints and semantic interoperability requirements. This document builds on D4.1 (Interoperable and Secure flow of Information) and D4.2 (Risk Identification Database & Risk Reduction Measures Database) to provide digital solution providers and water utilities with the tools and means to perform security assessments of the cyber-physical flow of information in strategic, tactical, and operational dimensions. It does so by targeting two different levels: first a risk guide to perform risk assessment of a given water system against cyber-physical threat is described, acting at the strategic and tactical decision level; in a second time, a set of recommendations for digital solutions, tailored made based on the output of D4.1 and D4.2, is presented. Given the comprehensive approach presented, from strategic to operational, this document should be of interest of the decision makers of all the levels in a water utility at least to create awareness on the relevance of increasing cyber security maturity. When it comes to the adoption of the approaches presented, the risk managers of water utilities are the target audience for the risk guide, while technology developers are the target audience of the set of recommendations. Furthermore, the set of security checks recommendations should also be of relevance for water utilities to be aware of the need of requiring security checks from suppliers when adopting new digital solutions to not increase the system vulnerability.

After the Appendix A which contains the Risk Guide, this document provides a set of security recommendations (gathered in Appendix B) that can be used on their own by technology providers and water utilities to both assess and mitigate risks, but also to develop secure solutions by design.

The risk guide and the security recommendations were tested in practice on two different solutions: the risk guide was tested on the interoperable Decision Support System (DSS) and real time control algorithms for stormwater management in Copenhagen, while the security recommendations, and in particular the black box security testing of IoT solutions was performed on the Low-cost temperature sensors for real-time combined sewer overflow (CSO) and flood monitoring.

Overall, this document describes a full process for secure integration of digital solutions within water utilities.

1. Introduction

Task 4.2 deals with “*cyber-physical security of flow of information in operational, tactical and strategic dimensions*”. The Task is structured in four sub-tasks (4.2.1 - 4.2.4) each of them aiming at supporting the digital water cities in performing risk management against cyber-physical threats, brought by the introduction of the DWC solutions, at different decision levels:

- 4.2.1 Strategic and tactical risk analysis based on the Risk Identification Database (RIDB)
- 4.2.2 Proposition of risk reduction measures on strategic and tactical level to reduce and mitigate risk events
- 4.2.3 System stress-testing against cyber-physical threats
- 4.2.4 Cyber-physical protection at operational level regarding DWC digital solutions

This deliverable is the result of the contribution of two sub-tasks: Task 4.2.3 and 4.2.4.

The contribution of Task 4.2.3 (complementing the work of Task 4.2.1 and Task 4.2.2) covers the strategic-tactical risk management, while the contribution of Task 4.2.4 covers the operational level.

At strategic-tactical level, the Task 4.2 methodology is inspired by the risk management procedure from the ISO Risk Management Framework (ISO 31000:2018) [1], including the steps of “Establishing the context”, “Risk identification”, “Risk analysis”, “Risk evaluation” and “Risk treatment” (Figure 1). Compatibility with this standard is key for the acceptance and interoperability of the proposed approaches with existing procedures in the water sector.

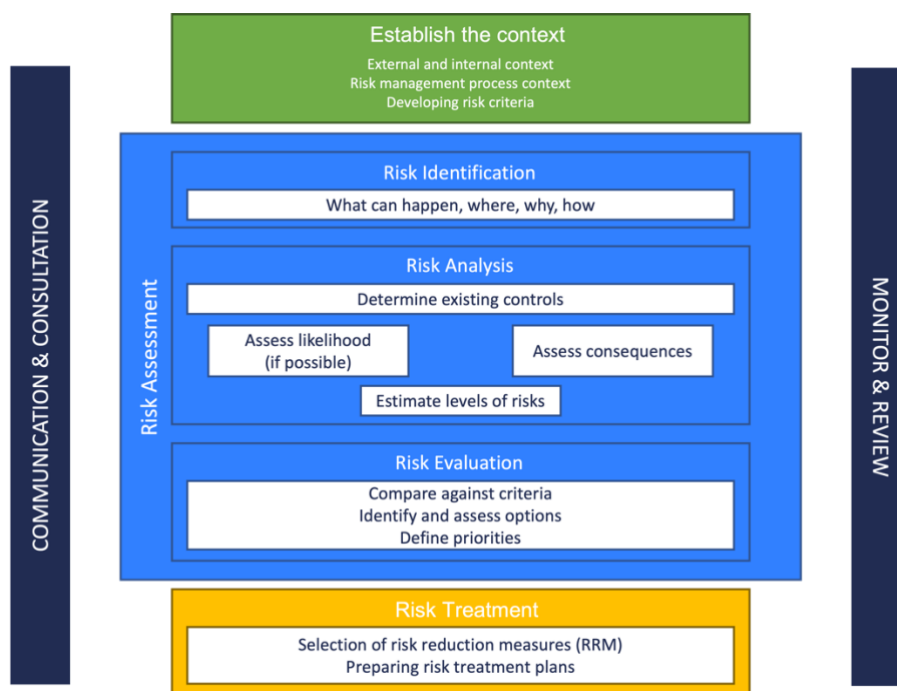


Figure 1. Risk management process as defined in ISO 31000:2018

D4.2 presented the DWC contribution to the steps “risk identification” and “risk treatment” through the creation of two databases, the RIDB (Task 4.2.1) and the RRMD (Task 4.2.2), as catalogues of risk

events and connected Risk Remediation Measures (RRMs) related to cyber and/or physical threats connected to the DWC solutions adopted by the DWC utilities [2, p. 2]. Task 4.2.3, in this deliverable, completes the work performed in D4.2 with instruction on how to perform the “risk analysis and evaluation” step. The methodology proposed is based on a concept of system-stress-testing. Finally, all the risk management steps are brought together in this deliverable in the form of a guide on how to complete a full risk management process against cyber and physical threats from risk identification to risk treatment. Each step is exemplified with a real use case, developed in collaboration with BIOFOS.

At operational level, D4.3 provides guidance for Digital Solutions to design products that can be integrated with DWC utilities to ensure the cyber-physical security of water and wastewater / stormwater systems in daily operation. This is achieved by task 4.2.4. building on the results presented in D4.1, where a baseline assessment of potential vulnerabilities of the IT and SCADA systems of the DWC cities are categorized for further follow up within this task. The outcomes of Task 4.2.4 consist of general recommendations to digital solutions, pointers to existing guidance documents to follow along with our own security checklists and set of requirements tailor made for the water sector. We demonstrated the value of performing security testing of the digital solutions before their integration by testing one of the real solutions.

Overall D4.3 aims at guiding the digital water cities to perform a complete risk management process against cyber-physical threats at strategic and tactical level, and at providing guidance for secure integration at operational level of the DWC solutions. Furthermore, this deliverable aims at raising awareness in the water sector on the topic of cyber security at strategic-tactical and operational level, bridging the gap between those two levels.

To achieve the set aims, this deliverable has been created with the following structure: we first provide the description of a Risk guide (reported in Appendix A), followed by our security recommendations for Digital Solutions. Together, our contributions help to get a full overview of the risks to Digital Solutions, thus helping to effectively manage it.

2. The strategic and tactical decision level: how to perform a risk management process against cyber-physical threats

2.1. Aim of the Risk Guide

A Risk Guide was developed with the aim of supporting water organisations in the implementation of risk management related to physical attacks and to emerging challenges connected to the digital solutions which support water processes. Indeed, together with physical threats, cyber-attacks are more and more frequently observed in water organisations, due also to the high increase of digital applications to improve the efficiency of the processes. Digital solutions can lead to a variety of cyber threats which are listed and described in Section 3 of this document.

Given the potential enormous damages which can derive from physical and cyber threats, a systematic approach to deal with the corresponding risks is presented in the Risk Guide reported in Appendix A of this document, which deepens the different steps of risk management, giving an overview of the approaches which can be implemented to protect water infrastructure against cyber-physical threats.

Within the DWC project, several digital solutions have been developed in the context of water organisations. Additionally, associated potential risk events and potential risk reduction measures

were identified by the project partners, and described within the Risk Identification Database (RIDB) and Risk Reduction Measure Database (RRMD) which are part of deliverable D4.2 [2, p. 2].

The Risk Guide aims also to show the applicability of the presented methodology by demonstrating it on one case study, addressing a risk event taken from the RIDB.

2.2. Content of the Risk Guide

In order to address the topic of risk management with a general approach which can be utilized by several types of organisations (not necessarily in the water sector), a procedure of Risk Management ISO 31000-2018 [1] was adopted as a structured framework which is represented in Figure 1. Risk management process as defined in ISO 31000:2018.

In particular, the Risk Guide focuses on the central boxes represented in Figure 1. Risk management process as defined in ISO 31000:2018, namely the green box “Establish the context” is addressed in the “Defining the context” section of the Risk Guide, the blue box of “Risk Assessment” is addressed throughout the “Risk Identification”, the “Risk Analysis” and the “Risk Evaluation” sections, and the orange box “Risk Treatment” is addressed with a homonym section of the guide. Each step of the risk management process is presented in the guide as:

- a. general description of the specific risk management phase with a list of possible methods which can be adopted by the organisations,
- b. exemplification through a use case where one or more of the proposed methods are applied for a case study. The use case has been produced in collaboration with project partners BIOFOS and DHI, developing the considered digital solution (DS13 "Web platform for integrated sewer and wastewater treatment plant control") within DWC.

The steps depicted in Figure 1 of communication and consultation, monitoring and review, and recording and reporting are not specifically addressed in the Risk Guide, however these boundary processes are strongly recommended throughout the document, especially in the “[Introduction](#)” (chapter A.1).

The introduction of the Risk Guide contains the key concepts of risk management and an overview of the application of the steps to the case study is presented.

In the first part of “[Defining the context](#)” in chapter A.2, several concepts are presented, such as the definition of risk as a combination of probabilities and consequences, Key Performance Indicator (KPI), risk criteria (e.g., numerical thresholds or risk matrix). In the second part of chapter A.2, the characterization of the analysed case-study system, the adopted risk criteria and the rationale of the study (e.g., the concept of dilution and related pollution) are reported.

In the first part of “[Risk Identification](#)” in chapter A.3), the different aspects needed to properly identify a risk event are listed and the concepts behind the RIDB, firstly developed within the STOP-IT project [3]; while in the second part of chapter A.3, the risk explorer of RIDB¹ is adopted for one selected risk event.

In the first part of the “[Risk Analysis](#)” in chapter A.4, qualitative, semi-qualitative and quantitative approaches are described together with a list of possible methods which can support the analysis of

¹ <https://risk-explorer.digital-water.city>

the risk, both in terms of consequences and probabilities. In the second part of chapter A.4, two sub-sections are reported, namely “Impact assessment” and “Probability assessment” since the risk is computed as the combination of consequences on the water system and probability of a successful cyber-attack. The stress-testing [4] of the selected water system is the adopted approach for the consequence assessment of the case study. Specifically, stress-testing a system implies multiple simulations of a variety of initial or boundary conditions, usually having the system under attack and/or under any performance lack of full capabilities because of process incidents or temporary performance degradation [5]. Concerning the probability part of the risk analysis, the tool InfraRisk-CP [6] is adopted to estimate the expected frequency of a successful attack for the case study. A set of questions were addressed to the partner BIOFOS, as the manager of the WasteWater Treatment Plant (WWTP) considered for the case study. In particular, the probability part was analysed starting from the answers provided by the involved organisation and its additional considerations about the constraints on the maximum and minimum values of the frequency of potential attacks and of the probability of success.

In the first part of the “[Risk Evaluation](#)” in chapter A.5, the principles about the usefulness of this phase are given, representing the ground for decision making on eventual risk mitigation measures to adopt, based on the application of the selected risk criteria. In the second part of chapter A.5, the selected risk criteria are applied to the case study and the need of risk treatment is discussed.

In the first part of the “[Risk Treatment](#)” in chapter A.6, the process of decision-making about the selected solution for risk mitigation is described and the concepts behind the RRMD (Risk Reduction Measure Database) [7] are described similarly to the RIDB; while in the second part of chapter A.6, the risk explorer of the RRMD is adopted to provide suggestion on the possible mitigation measure, followed by the discussion on whether to reduce consequences or probabilities with different solutions.

In fact, the risk connected to a certain contingency can be mitigated by barriers which effect either the causes of the event or the consequences of the event, as shown in the bow-tie diagram in Figure 2.

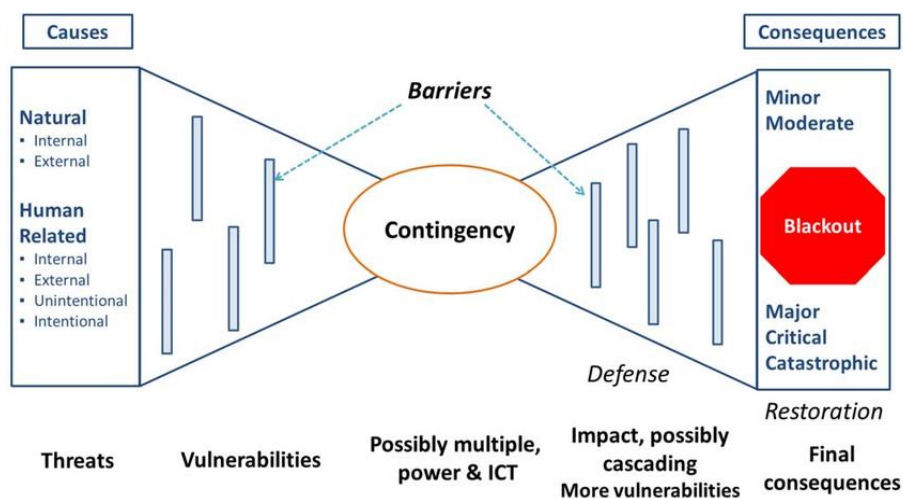


Figure 2. Example of bow-tie diagram related to the relationship between a contingency and its causes and consequences

2.3. Target users of the Risk Guide

The Risk Guide has been designed with the purpose of being applied by an organization which manages a critical infrastructure (CI) and wishes to improve its cyber-physical protection. Specifically, the Risk Guide should be used by a team of experts within the organisation who are able to go through the different steps of the proposed risk management approach. The team should have a clear overview of each responsibility around the involved assets, distinguishing a sub-group who is in charge of the overall organisation's strategies and another sub-group who is capable of providing knowledge about the system's processes and IT solutions.

The expertise on overall strategies should be particularly included in the “Defining the context” (when defining the risk criteria and the scope of the analysis), “Risk Identification” and “Risk Treatment” phases, while expertise on the physical and IT systems should be included within the steps of the “Risk Analysis” and “Risk Treatment” phases. Nevertheless, an intense communication between the different sub-groups of the team should be guaranteed all along the execution of the risk management steps, since the two aspects are strongly interconnected in all the mentioned steps.

In the case of water organisation, for the “Risk Analysis” it would be relevant to include an engineer or a group of engineers with qualified skills in hydraulic modelling and process engineering, since approaches such as Monte Carlo simulations or stress-testing procedures might likely be implemented on a digital twin of the water system. For the assessment of the probability of a successful attack, it is recommended to include different OT-technicians, with competences from the plant operations to plant maintenance, and an IT-technician with knowledge about the specific digital solution which could be hacked. Both types of skills are needed in the case of cyber-physical attacks. Setting up the team is part of the “Defining the context” step in ISO 31000:2018 [1].

2.4. Scope of use of the Risk Guide

it is recommended to use this guide at different times of the asset’s life (in terms of physical parts/solutions of the critical infrastructure and/or the components of a connected digital solutions which impact the processes of a water system or organisation), from the design to the implementation, to the decommissioning of the asset at its end of life.

In the design stage, the Risk Guide can support the creation of the solution from the very beginning of its conceptualization or during the testing phase, mitigating potential risks which are already identified a-priori.

In the implementation and operational stages, the Risk Guide becomes crucial since the effects of eventual attacks or incidents are real and have a tangible impact on the economics, on the society, and/or on the environment. Therefore, it is recommended to apply the guide periodically along the useful life of the asset which can be in the order of several decades, considering also that external or internal changes are likely to happen both on the cyber and on the physical layers.

In the decommissioning stage, the Risk Guide should be adopted, for instance, in order to ensure very low probability that a given dismissed digital solution represents an unintentional access point for hackers which might be capable to enter into the actual IT-system of the involved organisation.

3. The operational decision level: general security recommendations to digital solutions

This section aims to provide practical security recommendations to digital solutions' owners and to help raising security awareness within engineering teams by shedding light on the most common issues. The challenge in DWC was to give guidance to a wide variety of solutions, without that guidance being so generic that it would be useless.

As such, this section directly builds on the results of D4.1 which gave a baseline for security in DWC [8, p. 1] and on the Risk Identification and Risk Reduction Measures Databases (RIDB & RRMD) developed in D4.2 [1, p. 2] to derive a classification of the risks to DWC Digital Solutions. We also identify the different threats and how they relate to the expected security level.

As presented later in this section, digital solutions are built using a wide range of technologies, making it difficult to have one single and exhaustive list of recommendations. We chose instead not to reinvent the wheel and to present the digital solutions with a general approach to security, and with pointers to existing guidance, completed by our own recommendations, guidance and checklists when deemed necessary.

We start by building an understanding of the risks and threat landscape for digital solutions in DWC before looking at good security practices when it comes to development of such solutions. We end this section by using one of the digital solutions as a use case to demonstrate the value of security testing.

3.1. Risks to digital solutions in DWC

D4.1 "*Interoperable and secure flow of information*" described the different digital solutions in Digital Water City giving an overview of the technologies used and how they integrate (or not) with the water utilities in the cities. The solutions were then classified in three subsets, depending on their level of integration with the water utilities:

- **Standalone solutions:** These are solutions which are not interacting with any sensors or utilities. They can be Web or Mobile Applications, publicly available or requiring authentication.
- **Solution with "external" sensors:** These are the solutions that are gathering data from sensors "in the wild", using some long-range wireless technology (mobile networks, LoRa or Sigfox for instance) or manually gathered using most likely shorter-range wireless protocols, such as Bluetooth (Low Energy).
- **Solution with "internal" sensors:** These are the solutions that are gathering data from sensors that are placed in the water utilities (but not necessarily connected to their systems).

In D4.1, we also highlighted the fact that none of the digital solutions in DWC will be directly connected to a water utility: they will either not interact with the utility at all, or only share data with operators via an application (web or mobile). As such, we foresee two major risks for water utilities when integrating/using digital solutions:

- **Being led to take wrong operational decisions:** most of the digital solutions, while not interacting with the water utilities, provide crucial information to the operators of a water utility and are used as a decision support tool. The interoperable decision support system and real time control algorithms for stormwater management for instance, can be used by operators to predict the best maintenance window, thus optimising the process. If the information is erroneous though, it can lead to a release of untreated water in the

environment as a direct result of the wrong planning. Similarly, if the early warning system for bathing water quality reports an incorrect value so that bathing is authorized despite the quality being below the threshold, this can have disastrous consequences (both from a public health perspective, but also from a public relations one).

- **Supply-chain attacks:** as already mentioned in D4.1, water utilities being critical infrastructures, they must comply with strict regulations. It is safe to assume that they undergo regular audits and are supposed to be a difficult target to find an entry point to for attackers. This is the case for many industries and companies. To counter this, attackers might identify actors evolving around the main target and attack these companies instead, to later on leverage the trust in these companies as a mean to attack the real target. This is called a supply chain attack and has been used successfully by attackers in the past. The most damaging one in the past years is the SolarWinds Supply Chain Attack discovered in December 2020. The Network Management System used by hundreds of thousands of companies, was shipped with a malware, successfully hitting many high value targets such as the US Federal government [9]. In January 2021 for instance, vulnerabilities in Microsoft Exchange server were used to compromised hundreds of thousands of servers all around the world, including the European Banking Authority and the Norwegian Parliament [4], [6]. More recently, there has been several cases of malicious node packages being uploaded to npm² as an attempt to target companies such as Azure, Uber or Airbnb [13].

Motivations for attackers to attack digital solutions exist, independently of their level of integration with water utilities, and this justifies the need to ensure they are secure as well [14], [15]. Digital solutions vary a lot when it comes to the technologies used and to the services they provide, but it is possible to derive a high-level diagram of their architecture. Figure 3 presents such a diagram for DWC's solutions. Some solutions might be composed of a subset of the components presented here. The main components include:

- **Data sources:** Digital solutions usually rely on external data, which is then analysed and/or transformed to provide added value. This data can take various shapes: some solutions collect environmental data using IoT sensors deployed in the wild (water sources, sewer network, etc.), others use data from 3rd party services (weather or terrain information for instance) or even drones. The applications developed by a solution can themselves be considered as a data source; for instance, when collecting data from an off-line sensor using Bluetooth.
- **Solution infrastructure:** Most solutions rely on a backend infrastructure to operate their service. Infrastructure here refers to anything that supports services run by a digital solution and can consist of on-premises servers, cloud ones but also 3rd party services used as part of the data collection (Sigfox's network for instance). Network providers used for data collection are here considered as part of the solution infrastructure, contrary to the other external services described in the next point.
- **3rd party services:** These are the 3rd party services used by a solution to provide their own service, such as services providing SMS or email sending capabilities.

² npm is a package manager for the JavaScript programming language

- **Solution's services:** Solutions provide a service to water utilities/their users. This can be for instance an alert if the level of the E. Coli bacteria is too high in a water basin. A service can be exposed to the users in various forms, such as a web or mobile application, but also simply via an Application Programming Interface (API).
- **Users of the solution's services:**
 - o **Regular users:** Users of the service are for instance operators in a water utility in need to take operational decisions based on the information they receive from the digital solution's service. A good example could be operators visualising on its application that the level of E. Coli bacteria is higher than a given threshold in a water basin and deciding to forbid swimming there.
 - o **Machines:** If the service is exposed via an API, it might be used by another solution to develop something novel, or directly by a water utility to integrate it within their own system.

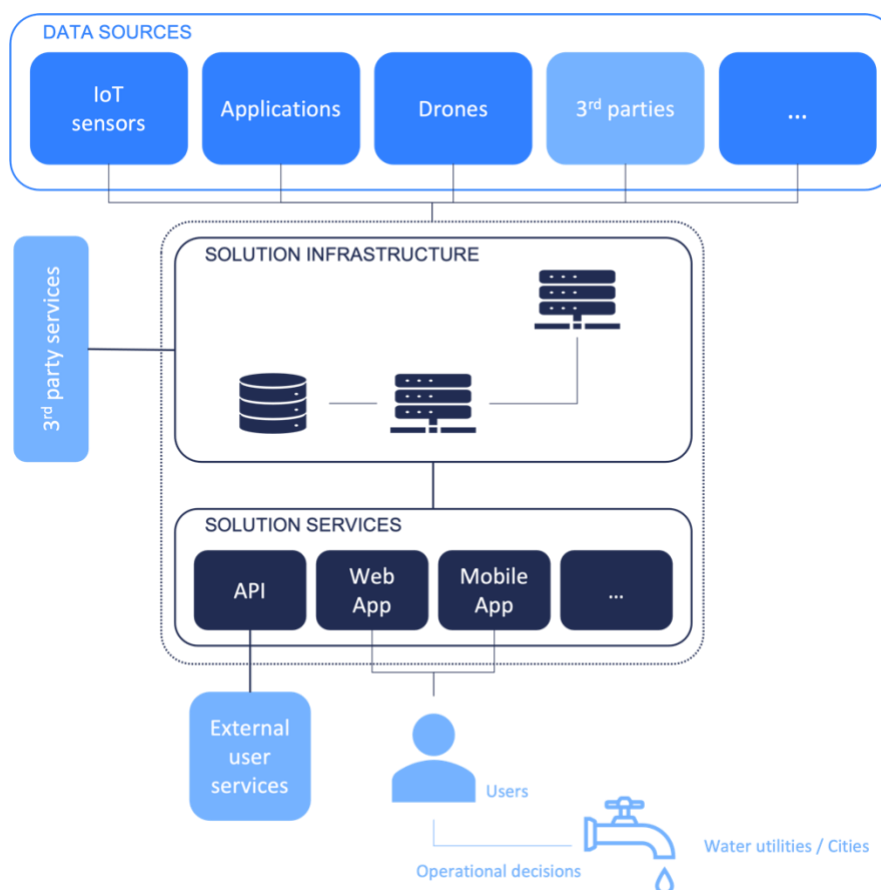


Figure 3 Generic Architecture Diagram for a Digital Solution in DWC

Digital solutions can be complex and might interact with several external actors to collect data, access services or to simply provide their own service to their users. Attackers thus get a wide choice of attack vectors when targeting a digital solution: IoT devices, applications, 3rd party services, etc.

In D4.2, “Risk Identification Database & Risk Reduction Measures Database” [2, p. 2], two databases gathering the generic risk events associated with the implementation of the digital solutions of DWC by the cities and the associated risk reduction measures³ were developed. Using the identified risk events, one can derive a classification of the attacks to the digital solutions in DWC. This classification is presented in Figure 4, and groups the attacks in 6 different classes which sum up the different types of attacks that can relate to a digital solution:

- **Attacks on IoT sensors:** As already explained above, IoT sensors are for many solutions a keystone as they constitute the source of data. Sensors are particularly vulnerable since they are often deployed in the wild (i.e., accessible by almost anyone), and are difficult and costly to secure (and thus have historically had poor security). While the data of the sensors themselves might not be very valuable to an attacker, being able to manipulate this data to trigger incorrect outputs by the services could have severe consequences. In addition, IoT sensors have in the past (and keep being so) already been compromised at scale to be integrated in botnets [16].
- **Attacks on infrastructure:** attacking the infrastructure is a way for the attackers to either disrupt the services (by for instance launching Denial of Services (DoS) attacks on it) or to gain unauthorized access to resources. This could also be a way for attackers to later use their access to a service to attack a water utility using and trusting this service.
- **Attacks on ML/AI:** compared to the other types of attack in this classification, attacks against Machine Learning (ML) and Artificial Intelligence (AI) are less known and appears a bit as newcomers. Attackers can for instance use specially crafted inputs to mislead the algorithms. Famous examples of such attacks include for instance cars being tricked into speeding by placing tape on speed signs.⁴ Other attacks consist of an attacker feeding incorrect data to the classifier, polluting the model in such way that its own data is later classified as good data, or on the contrary so that good data is inf act classified as incorrect. Finally, models have an intrinsic value, and an attacker might want to steal them.
- **Attacks on applications:** Applications (Web, Mobile, API, etc.) are usually exposed and if compromised, can lead to data leak, unauthorized access to resources and actions or allow for data manipulation and denial of service.
- **Human errors/failures:** while not being an attack per se, human error can lead to the same consequences. If a user is given access to data or actions he/she should not have access to, he/she could misuse it (intentionally or not) and effectively create a situation similar to an attack (for instance, a user could be given access to an alert system and trigger an alarm, leading operators to take decisions based on misleading data).
- **Social engineering:** Like human errors, a user could be tricked by a malicious person into performing harmful actions, potentially leading to dangerous consequences as well.

This classification, along with the threat landscape presented in the next section, is used as base for the general security recommendations for digital solutions’ owners.

³ The two databases are available on GitHub: <https://github.com/SINTEF-Infosec/dwc-risk-explorer> and can be explored dynamically at <https://risk-explorer.digital-water.city/>.

⁴ <https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/>

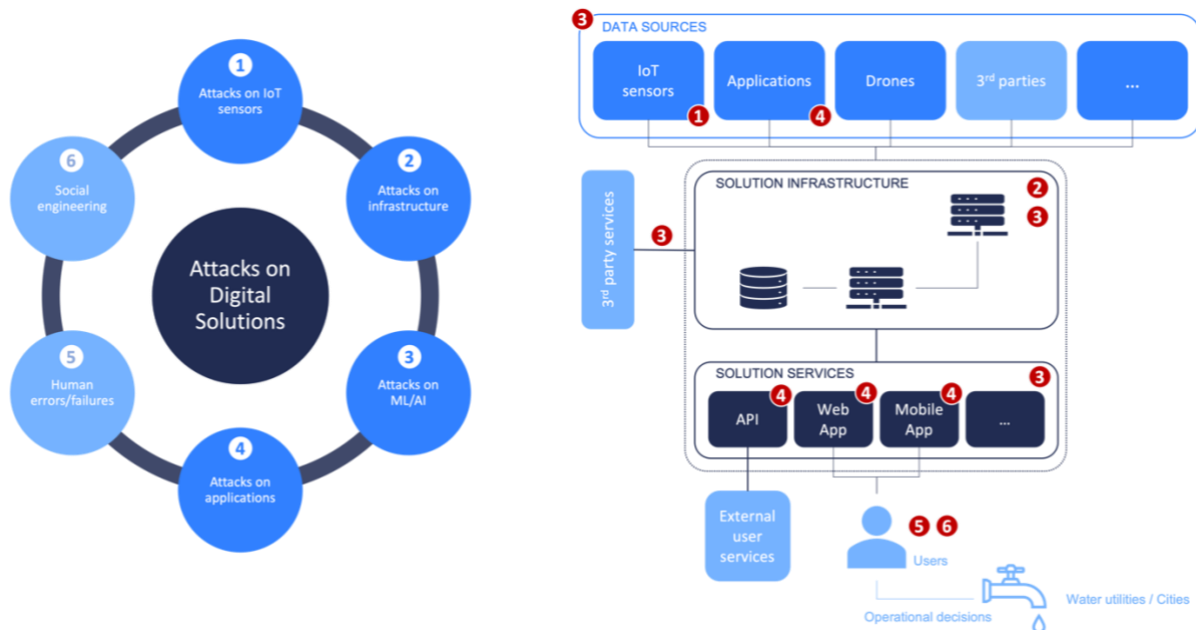


Figure 4 Classification of the attacks against DWC Digital Solutions

3.2. Threat landscape for the digital solutions

When securing a digital solution, it is important to take into consideration who and what one is defending against. Indeed, defending against a high-school student, running automated tools he found on the Internet, is not the same as defending against a state-sponsored threat actor that has unlimited resources. As presented by Weingart [17], we can classify attackers into three different classes based on their capabilities. While his classification is intended for physical security of embedded devices, it can be slightly adjusted the context of DWC. Our adapted classification is as follows:

- **Class I:** A clever outsider, who has limited knowledge about the system and a low budget and equipment. This could be a curious attacker that is targeting the system mostly for prestige and as a hobby, but also a “script kiddie”,⁵ dumbly following scripts and tutorials found on the Internet.
- **Class II:** A knowledgeable insider, who has advanced knowledge and/or specialized education and experience in the area. This category has access to sophisticated tools. Typically, this class corresponds to researchers.
- **Class III:** A funded organization categorized by its high budget and its ability to recruit class II attackers to attack the system. This corresponds to organized crime or to a government.

As mentioned in 3.1, attackers have incentives to target digital solutions to disrupt or gain access to water utilities. It is thus expected solutions should consider attackers from all three classes. However, for many solutions, defending effectively against state-sponsored actors is simply not feasible, as it

⁵ A script kiddie is person who uses existing computer scripts or codes to hack into computers, lacking the expertise to write their own.

would drastically increase the cost of their product, making it too expensive for any water utility to adopt. This is especially true for IoT based solutions, which rely on low-cost environmental sensors to collect data. Securing these devices while keeping the cost low is a challenge given today's state of the art in the area: the devices being deployed in the wild most of the time, they can be accessed by a malicious person and analysed not only from a software, but also from a hardware perspective. Microsoft's third immutable law of security states that *"If a bad guy has unrestricted physical access to your computer, it's not your computer anymore"* [18], and this holds even more true in the context of IoT devices. This, however, does not mean the solutions should give up on security, as they can still protect against class I and class II attackers, by for instance tackling the low hanging fruits in their product.

To secure a product against all classes of attackers, a change of paradigm is required: one must work with the assumption that parts of the solution will be broken/accessed by attackers (typically an IoT sensor) and ensure that the impact of this breach has no operational or financial consequences. There is no "one size fits all" scenario, and solutions must assess on their own where they stand, and how much effort is needed to secure their product. Indeed, a solution owner might be concerned by the Intellectual Property (IP) an attacker could get his hands on if he were to compromise an IoT device (models, algorithms, etc.) and thus choose to invest in more hardware security than for another solution that only measures environmental data to send them back to a backend infrastructure for processing.

Another way for solutions to see the problem is via cost: attackers, no matter which class they belong to, will go for the easiest and cheapest path that has the most impact. As such, ensuring the low hanging fruits are tackled will increase the cost and difficulty of an attack. Reducing the impact (by for instance ensuring proper segmentation) also contributes to attackers looking elsewhere. Since IoT is an easy target which can lead to high impact, we tried in this document to shed light on the risks linked to it by performing an assessment of one of the digital solutions in DWC (see 3.5), following the methodology presented in the next section.

Overall, the digital solutions' owners are advised to consider defending against attackers from all three classes. Defending effectively against all three within a reasonable budget being impossible, solutions should perform security risk analysis and weigh the cost and benefit of their different options.

3.3. Security testing methodology

For the digital solutions' owners to be able to weigh the cost and benefit of their different options, they first need to be aware of the risk magnitude. While security must be considered as part of the design process (*security by design*), having a way to ensure that there are no major security flaws in the product is extremely valuable. Indeed, even if the solution has been designed with security in mind, programming errors or misconfigurations are common and can cancel out those efforts. In this section we present different approaches to (practical) security testing of a solution.

3.3.1. Penetration testing

Penetration testing is defined by ENISA as *"the assessment of the security of a system against different types of attacks performed by an authorised security expert. The tester attempts to identify and exploit the system's vulnerabilities. The difference between a penetration test and an actual attack is that the former is done by a tester who has permission to assess the security of the system and expose its security weaknesses. In addition, the tester is given certain boundaries to operate and perform this task."* [19]

Penetration testing is usually divided into three subcategories, based on the level of knowledge the security expert is provided with at the start of the assessment: white, grey, and black box testing.

White box testing

In a white box testing configuration, the security expert performing the assessment is given full knowledge of the tested system (code, architecture documents and other useful information). The objectives of a white box testing are to simulate an insider's attack (for instance a previous employee who has had access to the code or even developed part of it). The advantage of a white box testing is that it might uncover bugs and vulnerabilities that can be difficult to identify by the other tests within a short timeframe. This benefit can also be seen as a disadvantage, as some of the uncovered vulnerabilities might not be "real vulnerabilities", in the sense that they cannot be exploited by an attacker (one can for instance imagine a function being flawed and leading to a buffer overflow when called with the wrong parameters; but if it is only called with constant values by the programmer, a malicious attacker will not be able to trigger it).

Grey box testing

Grey box testing is a mid-step between white box testing and black box testing: the security expert is given some information regarding the system targeted. This can be for instance indications on the technology used in the backend, available services (supposed to be non-accessible for instance) or some of the algorithms used. The pro of a grey box testing is that it combines some of the "real-world" advantages of the black box methodology (see below), while being more effective (the security expert does not lose time testing for things he could have known directly by looking at the documentation).

Black box testing

In a black box testing scenario, the security expert has very little or no previous knowledge of the targeted system or device. At a high level, the methodology consists of sending inputs to the system, the "black box", and to analyse the obtained outputs to deduce the internals of the target. Having made some guesses, the attacker can adjust her inputs to confirm her thoughts or to exploit the target. This is presented in Figure 5.



Figure 5 High level diagram of the Black Box Testing Methodology

The black box testing methodology has several advantages over the white and grey ones. Indeed, its primary objective is to test a system under real conditions, to emulate a real attack scenario. This means that such a test might catch errors made during the deployment of the system such as default passwords, misconfigurations in general or even the lack of security trainings of operators (weak passwords). This methodology also presents a low false positive ratio as the security expert can assess the risks associated with a vulnerability directly, i.e., if the vulnerability can be exploited or not. While black box testing can miss some vulnerabilities and should likely not be the first test performed, it is

an excellent way to assess how a system stands against attacks and to get an idea of the path an attacker would take to compromise the solution, and thus gives indications on how to tackle those potential low hanging fruits. It can be later be completed by a deeper assessment following a grey or white box approach.

3.3.2. Red team exercise

In a penetration test, there is no element of surprise, and the scope is limited to the system being assessed. A red team exercise goes one step further and includes physical penetration and social engineering: the objective of the attackers is to compromise a target by all possible means without being detected. This allows the company/organisation being tested to not only detect vulnerabilities in their system (like in a penetration test), but also to test their detection and response capabilities. In the context of DWC, red team exercises would be more targeted towards mature organisations such as water utilities.

3.4. Recommendations to digital solutions' owners

This section gathers our recommendations to the digital solutions' owners for a secure integration with the water utilities. Given that the solutions have little to no integration with the SCADA system (see D4.1), the recommendations come down to secure development of a digital solution.

We first provide general recommendations for security that apply all along the solution lifecycle (from design and implementation to operation) before covering the different attack categories defined in 3.1 and how to prevent them. As mentioned previously, digital solutions in DWC are diverse when it comes to the service they provide, the technologies they use and how they integrate with the utilities. As such, it not possible to cover all security measures in an exhaustive way. The main outcome of this task is thus a set of checklists that can be used by the solution providers if they deem their solution could be threatened by the covered attacks, as well as recommendations that solutions can use as a starting point to establish their security process.

3.4.1. General recommendations

Back in 2000, Bruce Schneier wrote *"Security is a process, not a product. Products provide some protection, but the only way to effectively do business in an insecure world is to put processes in place that recognize the inherent insecurity in the products. The trick is to reduce your risk of exposure regardless of the products or patches."* [20] This still stands strong, if not stronger, by today's standards.

The following recommendations aim to help build a security process when developing a digital solution. They have been grouped in three different categories covering the full lifecycle of a solution: its design, implementation, and operation. Those recommendations are generic and apply to any solution.

Designing

Know your enemy

The first step in designing a secure solution is to get an understanding of who and what to defend against. Companies might not always be aware that they can be targets, but as already mentioned, they can sometimes simply be used as a means to reach the primary target (supply chain attacks). As

such, companies must have an idea of the different profiles of the attackers, from a skilled hobbyist to a state sponsored hacker group or even an employee made unhappy by being denied a raise, or recently fired.

While it is important to know the profiles of the attackers, it is maybe even more important to understand how they operate or, in “technical terms”, what are their tactics, techniques and procedures (TTP) [21]. It is advised to keep up to date with threat intelligence provided by security companies such as Mnemonic or Trend Micro.

Overall, it is advised to acquire an adversarial mindset when developing a solution. Having an idea of what is possible, or simply what are the different attacks might not be an easy task when one has no security background. Frameworks like the Mitre ATT&CK⁶, a knowledge base of adversary tactics and techniques based on real world-observations, can help getting an overview of how attackers operate in the wild.

Define your security requirements

When building a new product, one of the first stages is to define a set of requirements for the new system. One usually focuses on *functional requirements* which define the different features of the products, but non-functional requirements such as security and reliability requirements also come into play. Some of those features might come from external sources, such as customers or regulations, and might at first seem in conflict with some of the functional requirements. One might then have to make trade-offs. It is often tempting to postpone dealing with security to instead focus on the functional requirements, but security is not optional for services in production, and a lack of it can lead to loss of business or worse, bankruptcy of the company [22]. Ensuring security by design might have a slight over-cost initially, but it pays out eventually.

Design for least privilege

The *principle of least privilege* states that “users and programs should only have the necessary privileges to complete their tasks.” [23] This limits the attack surface for attacks (internal attacks or compromise) and mistakes. In practice, this principle can be applied directly in the design by making each component/program do one thing well, and one thing only. This allows the application of the principle, by granting the necessary permissions to use a specific program (for instance an API endpoint). In addition to this, it is important to have good audit capabilities of the system to detect any malicious use of it, or simply to understand what happened in case of an incident, should an incident occur.

Design for a changing landscape

A product is usually designed with longevity in mind, and as such, the company is prone to see several changes in the threats to their product, but also in applicable regulations and expectations from their clients: those expectations are much likely to rise than to go down. In addition, a company/product might grow, and with this growth will likely come more security requirements. Past events have shown that companies failing to adapt, or having a poor initial design are likely to fail: some have even been forced to go bankrupt after a successful cyberattack on their system. This is for instance the case for “The Heritage Company” who had to shut down back in December 2019 after a ransomware attack,

⁶ <https://attack.mitre.org/>

they did not manage to recover from [24]. This highlights the need to design a product with these considerations in mind. In practice, this translates to building an infrastructure which makes changes easier: for instance, using containers helps keeping the application and its dependencies in a single package decoupled from the underlying OS, easing the update process (one can simply patch/update the image in the container registry, effectively helping to patch as part of the code deployment process). Another potential recommendation is to use microservices when suitable: splitting services in smaller units helps with scalability, maintenance, and security patching.

Design for resilience

Companies should strive for their solution to be resilient i.e., prepared to handle the unexpected. This can be a cyberattack or simply a legit increase of traffic for instance. Some strategies and mechanisms can help with this. Defence in depth is one of them. It consists in applying several defence mechanisms in a row, thus forcing an attacker to defeat all of them before reaching his goal. In the case of a web application, this could be for instance first a web application firewall, then access control on the API endpoints of the application, and perhaps containerization of the service. An attacker would then have to bypass all measures to get access to the host machine. In addition to defence in depth, it is recommended to have automated response mechanisms in place. This could for example be load shedding (a service replies with an error code if overloaded) or client throttling (increase of the response time) in the case of web service. Finally, even in the case of a compromised service, it is important to have mechanisms to reduce the impact, the most efficient being perhaps segmentation and strict application of the least privilege principle already mentioned above.

Additional considerations

While building security in one's product is important, one must sometime be pragmatic and consider the solution developed in its global context. Indeed, too much security can sometime have the opposite effect of what is expected. The best example for this is probably a password policy being too complex, which leads to users not able to remember their passwords, and them ending up writing those passwords down on a sticky note fixed on their screen. This happened for example to the French channel TV5 Monde, whose staff accidentally showed their passwords during an interview [25]. Similarly, security might impact the fluidity of operations, especially in a crisis. Facebook's outage in October 2021 is good proof of it: a configuration error on a BGP router disconnected Facebook's network from the Internet, and due to strict security policies, engineers faced difficulties to access the data centre, and later on to access the hardware itself to solve the issue.

Implementing

Ensure security by design

As already mentioned in the design section, security must be thought at the design stage, but it should also be the default option chosen when developing the solution. As an example, one can think about a solution exposing a service over HTTPS. While there could be reasons to allow the user to disable HTTPS and use HTTP, the default option should be to use HTTPS. Similarly, one must ensure that the

solution will fail securely, i.e., if a piece of code fails, the software fallbacks to a secure state. The following pseudo-code taken from OWASP⁷ exposes the problem clearly:

```
isAdmin = true;
try {
  codeWhichMayFail();
  isAdmin = isUserInRole("Administrator");
}
catch (Exception ex) {
  log.write(ex.toString());
}
```

If either the `codeWhichMayFail` or `isUserInRole` function fails, the user will be administrator by default.

Use frameworks

Whenever possible, one should not reinvent the wheel, but rather rely on existing frameworks. Writing applications that maintain security properties quickly become challenging as the project grows. Frameworks often provide all the necessary boilerplate to deal with common tasks regarding security (user authentication and authorization, logging, XSS and CSRF protections, etc.). Not having to re-create those basic blocks for every new solution helps reducing costs, but also helps with maintenance. Examples of such frameworks for web applications include Django, Laravel or Spring Boot. This recommendation also applies for embedded systems where embedded OS / Real Time Operating Systems provide developers with the means to implement things securely (ZephyrOS⁸ and FreeRTOS⁹ are such examples).

Be aware of common security pitfalls

During the implementation phase, it is recommended for engineers and developers to have common security pitfalls in mind. Initiatives such as the OWASP Top 10¹⁰ or the SANS Top 25¹¹ can provide food for thought for engineers to build upon. Checklists provided within DWC also aim to help building awareness regarding common security issues, not only in web applications but in the different areas identified in 3.1.

Test your solution

Solutions should test their solution as much as possible and using different approaches. To err is human, and even the most careful engineers make mistakes or forget about edge cases. In practice, this means that there are many ways for an untested solution to fail in the real world. Some of the recommended testing techniques include:

- *Unit testing*: breaking the software in a small testable unit that have no external dependencies. At that stage, external components such as databases are mocked. Most languages have frameworks to help with unit testing.

⁷ https://wiki.owasp.org/index.php/Security_by_Design_Principles#Security_architecture

⁸ <https://www.zephyrproject.org/>

⁹ <https://www.freertos.org/>

¹⁰ <https://owasp.org/www-project-top-ten/>

¹¹ <https://www.sans.org/top25-software-errors/>

- *Integration tests*: those tests take place one level above the unit tests and use real implementation of the databases and other components. These tests are about ensuring components integrate properly and securely together and that the system as whole behaves as expected.
- *Static code analysis*: as the name indicates, static code analysers inspect the source code of a program to detect potential errors or vulnerabilities. They can help catch issues before the code is built and deployed.
- *Dynamic code analysis*: this approach analyses the developed software in a dynamic manner to profile the performance, evaluate code coverage or security correctness. Dynamic analysis can identify several errors as well, such as race conditions, initialized memory being read, memory leaks or out-of-bounds memory accesses to quote a few.
- *Fuzzing*: fuzzing complements the above-mentioned techniques and consists in generating large amounts of inputs to the components being tested and analysing how it reacts to those inputs. It is particularly useful to detect bugs in parsers or protocol implementations. Fuzzing is even included by default in some languages such as Golang.¹²

Ensure secure deployment

It is best practice for new code added to the project to be reviewed before being pushed to production (or even checked in). Enforcing code review for every piece of code also helps protecting against malicious employees. Code review must then be mandatory, and one should not be able to opt out. This not only helps catching potential security flaws but also “regular” bugs and errors.

When deploying code, one should rely on automation as much as possible: automation removes the human from the loop, thus preventing mistakes and providing a consistent, repeatable process for building, testing and deploying software. [26]

Operating

Perform Penetration Testing and Red Team Exercises

It is important for solutions to have their solutions tested by performing both penetration testing and red team exercises. As already explained in section 3.3, those two methodologies are complementary and provide a good overview of the “real-life” security of the product. One could also imagine that water utilities might be interested in using Red Team Exercises to get an assessment not only of their technical security level, but also of their operational preparedness.

Prepare for vulnerabilities

New vulnerabilities are disclosed every day, either in a software used directly by a solution (OS, container technology, etc.) or in a library used to develop the solution. Vulnerabilities can also be discovered in the solution itself. As such, it is important for a solution to be prepared to handle vulnerabilities (either the ones impacting the solution, or the ones within the solution). This means having some sort of monitoring in place to be aware of the new vulnerabilities, but also a plan to patch, disseminate and communicate about vulnerabilities found in the solution with their users.

¹² <https://go.dev/doc/fuzz/>

Have a crisis management plan

While all the measures described previously in this section aim at preventing security breaches, one must plan for the worst, and as such for a security breach. If such an incident should happen, one must have a plan to deal with it in order to keep the system up and running in a secure and reliable manner if possible. This is a huge topic in itself, and we invite the solutions' owners to refer to the Google SRE book linked below.

While this whole section provided a wide range of information regarding security in the design, implementation and operation phases of digital solutions, its objective was to give an overview of the different activities to perform and have in mind, and as such, barely scratched the surface. The following resources were used to build this list and are recommended to dig deeper on the topic:

[OWASP-- Security by Design Principles](#)

[OWASP – Vulnerability Disclosure Cheat Sheet](#)

[Google SRE-- Building secure and reliable systems](#)

[OWASP SAMM v2](#)

[Security Engineering Book](#)

[ISO/IEC 27034](#)

3.4.2. Attacks on sensors/IoT

Attacks against IoT represents a fair part of the potential risks identified in the Risk Identification Database of DWC. IoT solutions in DWC are similar to other industries when it comes to the technologies used and to their design. As such, we find the same issues, and can also provide the same advice. We provide in appendix a checklist for IoT Security. The checklist questions are divided in a generic section and in 4 main areas that constitutes a high-level division of most IoT solutions:

- **Hardware:** the device itself (PCB design, debug ports, etc.).
- **Software:** the code running on the firmware (not the code running on the infrastructure).
- **Communication:** how is the data acquired and transmitted.
- **Infrastructure:** the backend infrastructure, servers, application, etc. Those questions overlap with the recommendations in 3.4.3 and 3.4.5.

This division highlights another important aspect of IoT solutions: they require a broad range of expertise (from hardware to backend going through embedded system development) and as a result teams are usually only experts on their specific topic and might work with assumptions on what other teams do, especially when it comes to security. It is thus vital for solutions that the teams have a good understanding of what level of security each team can provide and not to work with assumptions.

In addition to our checklist, existing guidelines such as the "[Baseline Security Recommendations for IoT in the context of Critical Information Infrastructure](#)" [27] from ENISA can be used to get a better understanding of the risk linked to IoT and of the countermeasures that can be applied.

3.4.3. Attacks on servers/cloud

Backend infrastructures of the solutions are the most at risk elements in the system when exposed to the Internet and must be protected accordingly. We provide in Appendix a *checklist for security in*

water process control networks as well as requirements for cloud security for critical infrastructures, derived from several best practices.

3.4.4. Attacks on ML/AI

While Machine Learning (ML) goes back to the 60s, little to no attention has been brought to ML security, until recently. Gary McGraw, Harold Figueroa, Victor Shepardson and Richie Bonett give a good overview of the risks to ML in *An architectural risk analysis of machine learning systems: Toward More Secure Machine Learning* [28]. This section summarized their work, but we recommend solutions working with Machine Learning to refer to the original document.

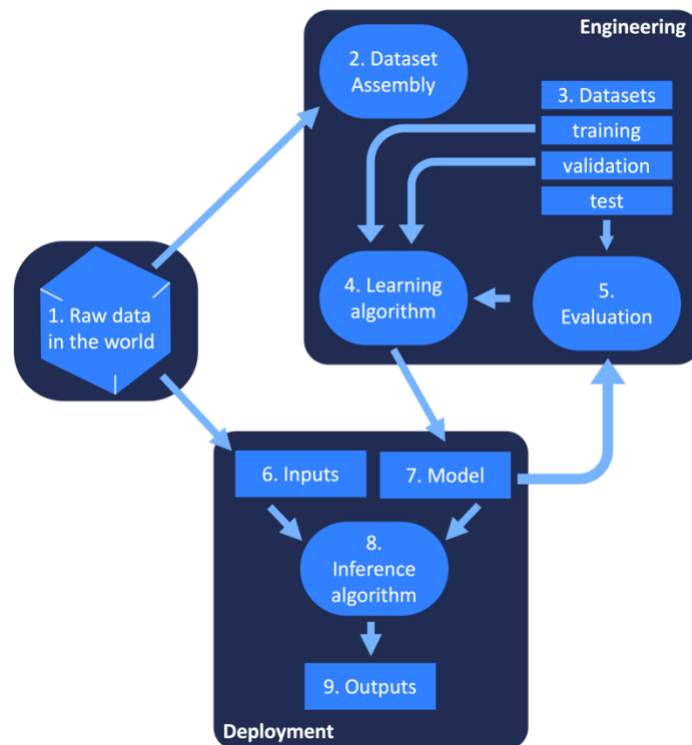


Figure 6 ML System architecture (adapted from [28])

Attacks on Machine Learning systems can be grouped in two categories: **manipulation attacks**, aiming at manipulating the system via manipulation of the data, inputs, or model; and **extraction attacks** aiming at getting information out of the system (the data, the inputs, or the model for instance).

3.4.5. Attacks on applications

The term applications carry here a broad meaning: an application can be code running on the backend infrastructure and processing data from external sources, or it can be a web application used by the utilities to access some data and perform operations, or it can be a mobile application used by technicians in the field to perform maintenance operations. Providing detailed recommendations for such a wide range of use cases covering many technologies is not possible, nor the objective of this deliverable. Instead, we suggest to both adhere to the general recommendations in 3.4.1 and to use methodologies for secure development and testing.

We recommend applications be developed following the secure development lifecycle (SDL) which contains the following phases: requirements, design, implementation, test, release. Activities and specific recommendations for those phases have already been covered in 3.4.1.

OWASP provide the [OWASP Application Security Verification Standard \(ASVS\)](#), targeting web applications, and the [OWASP Mobile Application Security Verification Standard \(MASVS\)](#), targeting mobile applications, which can be used to decide on a list of security requirements for secure development. They also provide two associated testing guides for both web and mobile applications, which can be used for security testing [29], [30].

3.4.6. Attacks on humans and human errors

To err is human. Back in January 2009, the message “*This site may harm your computer*” was displayed along with each search result made using Google search engine. The reason for this was simply a human error (a “/” was added to the list of malicious websites to flag, thus resulting in all websites being flagged) [31]. This shows that human errors impact everyone, even the bests.

There is very little difference between an employee making an error (for instance an engineer deploying the new solution with incorrect configuration) and an internal attacker doing so on purpose to allow access to himself later. Protecting a solution against human errors is thus equivalent to protecting against an internal attacker, and the recommendations discussed in 3.4.1 apply here too. For instance, designing for least privilege and requiring multi-authorization are way to mitigate or counter human errors.

Now, solutions can also be designed to effectively mitigate “legit” errors: the user interface should be clear and not allow confusion; users should be asked to validate their actions; etc. Finally, training the users on the tools is also important: both to prevent errors in the first place, but also to mitigate them if they happen.

For more information, one can look at ISO/IEC 27035 on incident response [32].

3.5. Security assessment of a DWC digital solution

While the objective of this document is to provide recommendations to digital solutions’ owners and cities on how to securely integrate the solutions to existing systems, we deemed important to demonstrate the usefulness of performing security testing on the digital solutions before integrating them. Given the scope and limited budget in the project, it was not possible to perform an extensive security assessment of all the digital solutions. It was thus decided to choose one solution that well represents digital solutions in DWC as a use case. Figure 3 represents the architecture of such a solution. After a discussion with ICRA about the solution *entitled “low-cost temperature sensors for real-time combined sewer overflow (CSO) and flood monitoring”*, it was agreed to use it as a use case. It uses low-cost sensors deployed in the wild and thus easily accessible. One of those IoT sensors was sent by IoTsens to SINTEF for security testing. SINTEF was also provided with a test account on the web application to verify findings during the assessment.

3.5.1. Methodology

The security assessment was performed following a black box testing methodology and the security researcher performing the assessment thus played the role of an external attacker with very little information on the system.

Our process can be split into five different tasks (see Figure 7). The very first one is the *Hardware Analysis*. Once the device is acquired, we started analysing its components, to know what the exposed interfaces are, debug interfaces, but also the chips that are on the board. To know that we had to open the device to access the Printed Circuit Board (PCB) to analyse it.

Knowing the components and available interfaces, we then started looking for documentation such as datasheets, Request for Comments (RFC) or any other relevant information about the device. The goal of that second step is to understand the overall system and come up with some first hypotheses about it.

From those hypotheses, we then came up with testing scenarios to be performed on the device. Those scenarios have two possible outcomes: either a success or a failure. However, the success or the failure of a testing scenario is determined by the expected result, which in reality means that even failure brings us new information about the system.

Indeed, the results of a specific scenario need to be interpreted. This interpretation is called a finding. Those findings are then used to look for new documentation and/or infer new hypotheses about the device. For example, if the hypothesis is *"the device has debug ports exposed on the Universal Asynchronous Receiver-Transmitter (UART) pins and is providing the attacker with a shell when connected to it"*, then the testing of that particular scenario will lead to either a confirmation or a rejection of the hypothesis. The interpretation is here quite easy as it is the hypothesis itself. This finding can then be reinjected in the documentation phase to infer new testing hypotheses like *"the attacker is given a root shell when connecting to the UART console"* or *"the attacker can access the filesystem when connecting to the UART console"*.

Those findings are finally gathered to be reported. The reporting step is the one where the device is considered back in its whole ecosystem. That means the findings are interpreted again, but that time with regards to different metrics. In the case of the above example, one can wonder what is the impact of the attacker having access to the filesystem on an IoT device? Linked to other findings such as *"The data is stored in cleartext on the device"* or *"The data is not stored on the device"* the impact and thus the interpretation can be very different.

Those steps can be mapped with the Open-Source Security Testing Methodology Manual (OSSTMM) [33] which is widely used to assess security of IT systems. Indeed, the first two steps (*hardware analysis* and *documentation*) correspond to the *information gathering* (or approach) phase in the OSSTMM. The *contact phase* is then used followed by the *exploitation phase*, which are here mapped with the *testing* and *findings phases*. In the OSSTMM, the information gathered during the first phase along with the one gathered directly by the *contact phase* is then used to exploit the system and gain access. In our process, the information required to exploit, and gain access comes from previous testing. Finally comes the *reporting phase*. In the OSSTMM, one more phase is sometimes used depending on the engagement: the *persistence* one. In our case, *persistence* is studied as a hypothesis which is then tested and reported as any other findings.

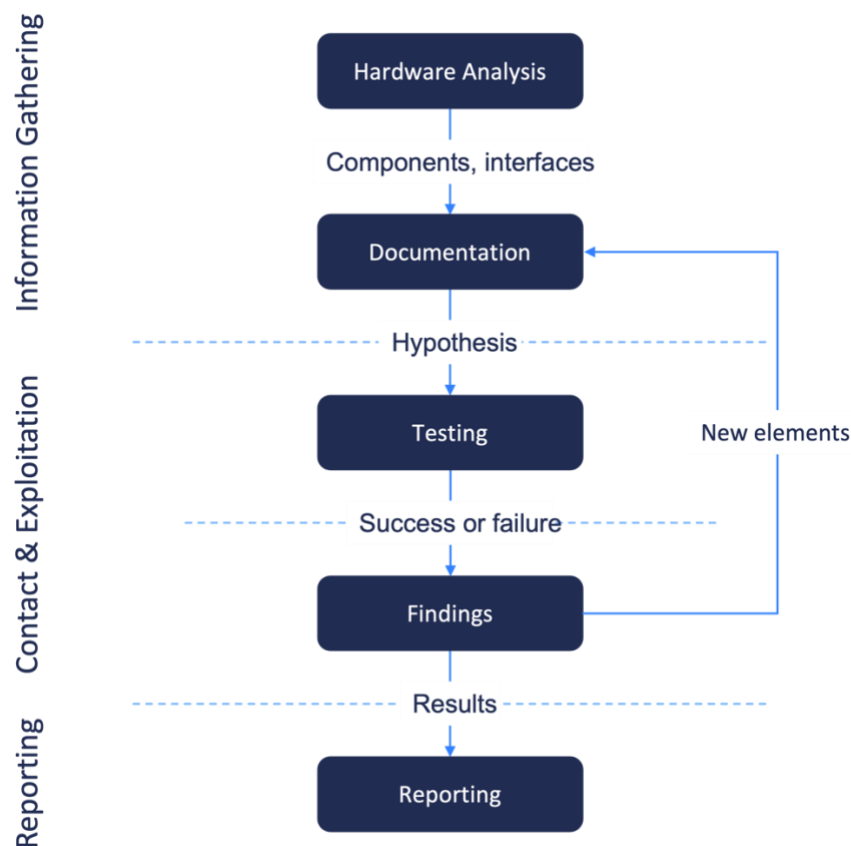


Figure 7 Black Box Methodology iterative cycles used during our assessment

3.5.2. Scope

Our starting assumption is that attackers want to use the digital solution to influence operational decisions taken by the water plants using its data. As such, we assumed that the attackers stole one of the sensors installed in the sewer to analyse it and see what they could find. We also assumed that the device delivered to us was in a production mode. It was however brought to our attention that the casing in which we received the device is not the one used in the wild, and as such, we did not comment on the casing itself.

The scope was limited to the device itself. It was configured to use the mobile network, and it was also agreed that the Long Range (LoRa) interface would be out of scope for this assessment. The device also presents interfaces for configuration (and likely updates), but those were deemed out of scope for this assessment as well for time and budget reason.

It was also agreed that interactions with the backend server was allowed as well as sending fake data if feasible (this was one of our objectives). The server being used for other activities than the security test, potential attacks that could have disrupted the service were not carried out but the concern was raised with the engineering team.

3.5.3. Results

This document being public, the findings from the security assessment have been reported separately to IoTsens via a confidential report. An additional meeting was also conducted with the engineering team to discuss potential questions from the company and shed light on some of our interrogations.

According to the results obtained in the report, from IoTsens a series of internal meetings have been held with all the personnel involved in the development of the solution to plan the corrective tasks. At the level of firmware programming, some changes have already been defined in its programming that will be implemented in the following updates of the equipment. On the other hand, at the hardware level, some actions that can be carried out in the next production phase have been studied with the design team. Finally, at software level, a more in-depth study plan has been established internally that will be carried out with the company's IT department.

3.5.4. Exploitation

While the findings of the assessment are confidential, we used the experience gained and the observations made during the assessment to expand and validate the security checklist for IoT presented in 3.4.2. This assessment was also used to validate and expand the Risk Identification Database and Risk Reduction Measure Database developed in Task 4.2.2. In particular, the following risks have been added¹³:

- Event 69: *External Attacker generates a cyber threat causing a Denial of Service of the Database of the Low-cost temperature sensors for real-time combined sewer overflow (CSO) and flood monitoring which affects CSOs data quality and might lead to a Financial issue.*
- Event 70: *External Attacker generates a cyber threat causing a Denial of Service of the Database of the Low-cost temperature sensors for real-time combined sewer overflow (CSO) and flood monitoring which affects CSOs data quality and might lead to a Quality issue.*

The following risk reduction measure has been added:

- M74: *Service throttling*

4. Conclusion

This deliverable provides the means to develop secure digital solutions within DWC. We show that while most of the digital solutions will not be directly connected to the water utilities, an attack on them can still have operational consequences. We demonstrate the feasibility of such attacks using both the risk guide (to increase preparedness at a more strategic-tactical level) and security testing on two different digital solutions (to increase security awareness at an operational level). We also show that both tools are complementary and provide a different overview of the risks to the solutions. Finally, we developed a list of security recommendations to be followed by digital solutions' owners when developing new products. We also foresee those recommendations to be used by water utilities to dictate security requirements for digital solutions' owners. While developed within DWC, most of the recommendations can be extended to other critical infrastructures.

¹³ <https://github.com/SINTEF-Infosec/dwc-risk-explorer/pull/8>

References

- [1] 14:00-17:00, 'ISO 31000:2018', *ISO*.
<https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/56/65694.html>
(accessed Apr. 28, 2022).
- [2] G. Bour, I. Selseth, M. Jaatun, and R. Ugarelli, 'D4.2: Risk Identification Database & Risk Reduction Measures Database', Nov. 2021, doi: 10.5281/zenodo.6497050.
- [3] A. Ostfeld *et al.*, 'D3.2 Risk Identification Database (RIDB), STOP-IT (2018)'.
- [4] D. Nikolopoulos, G. Moraitis, D. Bouziotas, A. Lykou, G. Karavokiros, and C. Makropoulos, 'Cyber-physical stress-testing platform for water distribution networks', *J. Environ. Eng.*, vol. 146, no. 7, p. 04020061, 2020.
- [5] C. Makropoulos *et al.*, 'D4.2: Risk Analysis and Evaluation Toolkit (RAET), STOP-IT (2019)'.
- [6] I. Selseth and R. Ugarelli, 'InfraRisk CP – User's guide, STOP-IT (2020)'.
- [7] H.-J. Mälzer, F. Vollmer, and A. Corchero, 'Risk Reduction Measures Database (RRMD)', STOP-IT deliverable D4.3, 2019.
- [8] H. Schwarzmüller, A. Vennesland, P. H. Haro, and G. Bour, 'D4.1: Interoperable and Secure Flow of Information - Cyber-physical Sphere and Interoperability Aspects in the Utilities Regarding the DWC Solutions', Mar. 2021, doi: 10.5281/zenodo.6497313.
- [9] 'What You Need To Know About the SolarWinds Supply-Chain Attack | SANS Institute'.
<https://www.sans.org/blog/what-you-need-to-know-about-the-solarwinds-supply-chain-attack/>
(accessed May 02, 2022).
- [10] 'At Least 30,000 U.S. Organizations Newly Hacked Via Holes in Microsoft's Email Software – Krebs on Security'. <https://krebsonsecurity.com/2021/03/at-least-30000-u-s-organizations-newly-hacked-via-holes-in-microsofts-email-software/> (accessed May 02, 2022).
- [11] 'Norway's parliament hit by new hack attack', *Reuters*, Mar. 10, 2021. Accessed: May 02, 2022. [Online]. Available: <https://www.reuters.com/article/us-norway-cyber-idUSKBN2B21TX>
- [12] 'European Banking Authority hit by Microsoft Exchange hack - BBC News'.
<https://www.bbc.com/news/technology-56321567> (accessed May 02, 2022).
- [13] 'A Large-Scale Supply Chain Attack Distributed Over 800 Malicious NPM Packages', *The Hacker News*. <https://thehackernews.com/2022/03/a-threat-actor-dubbed-red-lili-has-been.html>
(accessed May 02, 2022).
- [14] 'Governments need to reassess security infrastructures | Orange Business Services'.
<http://www.orange-business.com/en/magazine/new-generation-critical-infrastructures-secure>
(accessed May 02, 2022).
- [15] 'Clear the "air gap" myth to evade cyber threats - Securing critical infrastructure in the digital world', *Nokia*. <https://www.nokia.com/networks/insights/critical-infrastructure-enterprise-security/>
(accessed May 02, 2022).
- [16] G. Kambourakis, C. Koliass, and A. Stavrou, 'The Mirai botnet and the IoT Zombie Armies', in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Oct. 2017, pp. 267–272. doi: 10.1109/MILCOM.2017.8170867.

- [17] S. H. Weingart, 'Physical Security Devices for Computer Subsystems: A Survey of Attacks and Defences', in *Proceedings of the Second International Workshop on Cryptographic Hardware and Embedded Systems*, Berlin, Heidelberg, Aug. 2000, pp. 302–317.
- [18] Microsoft, 'Ten Immutable Laws Of Security (Version 2.0)'. <https://docs.microsoft.com/en-us/archive/blogs/rhalebher/ten-immutable-laws-of-security-version-2-0> (accessed Apr. 03, 2022).
- [19] 'Vulnerabilities and Exploits', *ENISA*. <https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/vulnerabilities-and-exploits> (accessed Apr. 03, 2022).
- [20] B. Schneier, 'Essays: The Process of Security - Schneier on Security'. https://www.schneier.com/essays/archives/2000/04/the_process_of_secur.html (accessed Apr. 03, 2022).
- [21] C. Johnson, M. Badger, D. Waltermire, J. Snyder, and C. Skorupka, 'Guide to Cyber Threat Information Sharing', National Institute of Standards and Technology, NIST Special Publication (SP) 800-150, Oct. 2016. doi: 10.6028/NIST.SP.800-150.
- [22] 'Ransomware Attack Topples Telemarketing Firm, Leaving Hundreds Jobless | Threatpost'. <https://threatpost.com/ransomware-attack-topples-telemarketing-firm/151530/> (accessed May 29, 2022).
- [23] NIST CRSC, 'Principle of Least Privilege'. https://csrc.nist.gov/glossary/term/principle_of_least_privilege (accessed Apr. 04, 2022).
- [24] 'Company shuts down because of ransomware, leaves 300 without jobs just before holidays | ZDNet'. <https://www.zdnet.com/article/company-shuts-down-because-of-ransomware-leaves-300-without-jobs-just-before-holidays/> (accessed May 29, 2022).
- [25] 'TV station gets hacked. And then broadcasts its passwords in a report about the hack', *The Independent*, Apr. 12, 2015. <https://www.independent.co.uk/tech/tv5monde-hack-staff-accidentally-show-passwords-in-report-about-huge-cyberattack-10168475.html> (accessed May 29, 2022).
- [26] H. Adkins, B. Beyer, P. Blankinship, A. Oprea, P. Lewandowski, and A. Stubblefield, *Building Secure and Reliable Systems: Best Practices for Designing, Implementing, and Maintaining Systems*. O'Reilly Media, 2020. [Online]. Available: <https://books.google.no/books?id=Kn7UxwEACAAJ>
- [27] 'Baseline Security Recommendations for IoT', *ENISA*. <https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot> (accessed May 29, 2022).
- [28] G. McGraw, H. Figueroa, V. Shepardson, and R. Bonett, 'An architectural risk analysis of machine learning systems: Toward more secure machine learning', *Berryville Inst. Mach. Learn. Clarke Cty. VA Accessed Mar*, vol. 23, 2020.
- [29] 'OWASP Web Security Testing Guide | OWASP Foundation'. <https://owasp.org/www-project-web-security-testing-guide/> (accessed Apr. 25, 2022).
- [30] 'OWASP Mobile Security Testing Guide | OWASP Foundation'. <https://owasp.org/www-project-mobile-security-testing-guide/> (accessed Apr. 25, 2022).
- [31] 'Official Google Blog: "This site may harm your computer" on every search result?!?!' <https://googleblog.blogspot.com/2009/01/this-site-may-harm-your-computer-on.html> (accessed May 29, 2022).

- [32] 14:00-17:00, 'ISO/IEC 27035-2:2016', ISO. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/20/62071.html> (accessed May 29, 2022).
- [33] ISECOM, 'OSSTMM'. <https://www.isecom.org/OSSTMM.3.pdf> (accessed Apr. 03, 2022).
- [34] I. A. Tøndel, M. G. Jaatun, and J. Røstum, 'IKT og sikkerhet i VA-sektoren: Hva kan gå galt?', 265-269, 2013, Accessed: Apr. 21, 2022. [Online]. Available: <https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2583766>
- [35] M. G. Jaatun, J. Røstum, S. Petersen, and R. M. Ugarelli, 'Security Checklists: A Compliance Alibi, or a Useful Tool for Water Network Operators?', 872-876, 2014, doi: 10.1016/j.proeng.2014.02.096.
- [36] K. Bernsmed, P. H. Meland, and M. G. Jaatun, *Cloud Security Requirements - A checklist with security and privacy requirements for public cloud services*. 2015. Accessed: Apr. 21, 2022. [Online]. Available: <https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2378820>
- [37] 'A 238 Informasjonssikkerhet og skybaserte tjenester for vannbransjen (kun digital) | Norsk Vanns Kompetanseweb'. <https://va-kompetanse.no/butikk/a-238-informasjonsikkerhet-og-skybaserte-tjenester-for-vannbransjen-kun-digital/> (accessed Apr. 21, 2022).

Appendix A – Risk Management Guide for Cyber-Physical Attacks in Water Systems

A.1. Introduction

Aim of the document

This document is conceived as a step-by-step procedure to guide organizations through the framework of Risk Management to deal with cyber-physical attacks in critical infrastructures such as urban water systems.

Scope and intended use

The Risk Guide has been designed with the purpose of being applied by an organization which manages a critical infrastructure (CI) and wishes to improve its cyber-physical protection. Specifically, the Risk Guide should be used by a team of experts within the organisation which are able to go through the different steps of the proposed risk management approach. The team should have a clear overview of each responsibility around the involved assets, distinguishing a sub-group who is in charge of the overall organisation's strategies and another sub-group who is capable of providing knowledge about the system's processes and IT solutions.

By following the approach presented in this guide, supported by the exemplification through a use case, the users can perform similar assessments, adapted to the specific context, by following the procedure proposed. The use case consists of a wastewater treatment plant (WWTP) with given characteristics under cyber-attack. The risk management process described by the ISO 31000-2018, in the form of specific steps, is adopted and adapted to the risk objective of protection against cyber and physical attacks. For each step, guidance is provided both in terms of theoretical approaches and modeling solutions, where relevant reference to solutions developed in other H2020 projects, i.e., the [STOP-IT project](#), are provided.

Project reference

This guide is part of the deliverable D4.3 of the H2020 Project Digital Water City (DWC) – Grant Agreement Nr. 820954. It has been developed as joint collaboration between the Partners SINTEF, BIOFOS and DHI. One of the Digital Solutions (DS) developed within the project (DS13 "Web platform for integrated sewer and wastewater treatment plant control") by BIOFOS and DHI, was selected as the target of a cyber-attack, taking into account a specific threat: the spoofing of the web interface where forecast data are visualized by operators leads to a wrong flow prediction and thus to a wrong maintenance schedule.

The risk management process

The ongoing digitalization in the water sector brings opportunities to increase the efficiency of processes, but also new challenges related to potential cyber-attacks on the water systems.

To manage risk events due to cyber and/or physical threats in the water systems systematic framework should be adopted. The methodology described in this guide is consistent with a Risk Management ISO (ISO 31000-2018), thus covering the following steps:

- a) [Context description](#)
- b) [Risk Identification](#)
- c) [Risk Analysis](#)
- d) [Risk Evaluation](#)
- e) [Risk Treatment](#)
- f) Monitoring and review
- g) Communication

The steps require continuous iterations since the Risk Management framework should not be intended just as a sequential process, as shown in Figure A.1. First, when managing risk, water organizations should define the scope of the activities related to the specific risks under investigation. Moreover, water organizations should specify risk criteria and KPIs which reflect their objectives and values, in order to define the amount and type of risk that they may or may not accept. While implementing the first five steps, continuous activities of communication and consultation, monitoring and review, and recording and reporting should be performed. The overall process of Risk Identification, Risk Analysis and Risk Evaluation is defined as Risk Assessment. Risk identification consists on identifying potential risk events to be further analysed in the following steps. Risk analysis consists in assessing the magnitude of risk, by estimating the probability of the selected event to happen and the consequences created. Risk evaluation involves comparing the magnitude of risk estimated during the risk analysis with set risk criteria defining the level of risk that is acceptable, tolerable, or not acceptable. Risk events can therefore be ranked in terms of severity. The result of this step is used to make decisions about future actions on risks events that need to be treated with the adoption of adequate risk reduction measures.

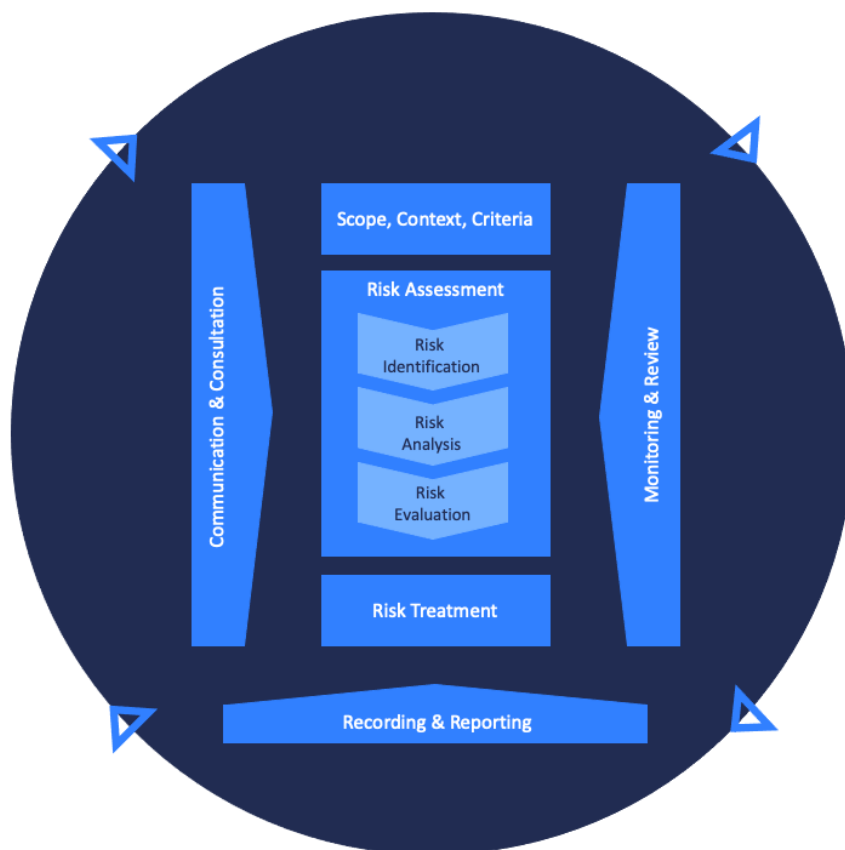


Figure A.1. The Risk Management framework (ref: ISO 31000:2018)

Each step of the framework of Risk management is explained to the users and then demonstrated through the application of relevant methods through the use case example to show how the different steps can be implemented. The features of the adopted case study are listed in Table A.1.

Table A.1. Characterization of the features of the case study

Risk management steps	Risk management steps applied to the example
<p>Defining the context</p>	<p>Scope: Improvement of scheduled maintenance at WWTP in Copenhagen</p> <p>System: Interface of a Web Application (DS 13) to visualize the inflow 48 hours forecast at the WWTP</p> <p>Risk criterion: Yearly untreated volume (KPI) < Target value</p>

<p>Risk identification</p>	<p>Risk event identified: Internal attacker generates a cyber threat causing a Spoofing of the Web Application of the DS 13 Interface which affects Sewers or Wastewater treatment plant and might lead to a Quantity issue</p>
<p>Risk analysis</p>	<p>Consequence: KPI computation through a stress testing procedure, i.e., simulation of multiple scenarios (considering as input the historical timeseries of water inflow at the entrance of the analyzed WWTP) under maintenance and under a Spoofing of the Web Application of the DS 13 Interface.</p> <hr/> <p>Probability: Estimation of probabilities of a successful cyber-attack performed by a semi-qualitative approach, based on expert judgements, on specific questions (Q&A approach) developed in the STOP-IT project (InfraRisk tool).</p>
<p>Risk evaluation</p>	<p>Comparison of results and risk criterion: According to the adopted pre-defined method, the estimated risk (given by the combination of consequence and probability) is compared with the target value, defining the level of risk.</p>
<p>Risk treatment</p>	<p>Risk mitigation measures: based on the Risk Reduction Measures Database (RRMD) - where several risk reduction measures are gathered and associated to related risks events of the RIDB - relevant measures are discussed for addressing and reducing the identified risk.</p>

A.2. Defining the context

Defining the context within ISO framework. The aim of the step “defining the context” is to set the goal/scope to perform the Risk Management Process (RMP) (for this guide the scope is related to protection of a given system against cyber – physical threats), to describe the system of interest, to describe the external and internal conditions that can influence the risk management steps, to define responsibilities and, remarkably, to set the risk criteria against which the risk will be assessed. RMP

should be aligned with the utility objectives and strategies and should target the specific risk affecting the achievement of those objectives. Thus, defining the context is a necessary step for risk management, which concerns the identification of all external and internal variables to be considered when managing risk.

Understanding the framework in which the water utility functions is necessary to assess what risks there may be, as well as what effect they could have. The system under investigation should be well defined and potential threats should be mentioned. As mentioned, establishing the context also includes the need to define risk criteria. The third step of the risk assessment process involves evaluating the risk against the organization’s own criteria, including regulatory requirements and stakeholder expectations. These aspects need to be defined at an early stage, including the types of consequences which are relevant and how they will be measured, the basis for considering likelihood and how risk levels will be determined and evaluated.

Specifically, in the risk analysis phase, the consequences assessed in terms of specific dimensions of impacts (economic, reduction of service, environmental, etc.), could be quantified, when possible, as Key Performance Indicators (KPI) and then combined with the estimation of the probability of a successful cyber-attack. For instance, if the probability of having a successful attack is 10% per year, the KPI value obtained as consequence within the impact assessment of the risk analysis should be multiplied by 0.1, so that the resulting KPI value represents the risk, given the following expression.

$$\text{Risk} = \text{Probability} \times \text{Consequence}$$

As a possible risk criterion, the computed risk in terms of the selected KPI may be compared with a target value so that the risk manager can decide whether to stop or to proceed with risk mitigation measures.

KPI ≤ Target Value → Risk Mitigation measures are **not** needed

KPI > Target Value → Risk Mitigation measures are needed

The risk criteria might also cover more shaded areas of risk severity characterization, not limited to “high” and “low” risk levels, but also cover intermediate zones of “medium risk”. The definition of the ranges of combinations of probabilities and consequences reflecting different levels of risk severity are often represented in the form of risk matrices

As an example, a scale of probabilities could be given by:

- Very unlikely (VU): Less than once per 100 year ($p < 0.01$)
- Remote (R): Once per 10-100 year ($0.01 \leq p < 0.1$)
- Occasional (O): Once per 1-10 year ($0.1 \leq p < 1$)
- Probable (P): Once to 12 times a year ($1 \leq p < 12$)
- Frequent (F): More than once a month ($p > 12$)

In terms of consequences, a scale should be again defined with respect to the selected KPI for the considered risk (note that the KPI values at this stage are not representing risk, but consequences).

- Very Low (VL): $KPI < 1$
- Low (L): $1 \leq KPI < 3$
- Medium (M): $3 \leq KPI < 10$
- High (H): $10 \leq KPI < 30$

- Very High (VH): $KPI \geq 30$

By combining the mentioned categories of probability (P) and consequence (C), an example of risk matrix is shown in Table A.2. The organization can assign different level of risk, color-coded in Table A.2 for different combinations of P and C.

Table A.2. An example of risk matrix with 5 levels of Risk

Probability (P)	F	(0,4)	(1,4)	(2,4)	(3,4)	(4,4)
	P	(0,3)	(1,3)	(2,3)	(3,3)	(4,3)
	O	(0,2)	(1,2)	(2,2)	(3,2)	(4,2)
	R	(0,1)	(1,1)	(2,1)	(3,1)	(4,1)
	VU	(0,0)	(1,0)	(2,0)	(3,0)	(4,0)
	VL	L	M	H	VH	
	Consequence (C)					

For instance, the given risk event may fall into the grey, green and yellow areas with low, acceptable or tolerable risk or into the orange and red areas with high or very high-risk level, so that mitigation measures are triggered accordingly.

Defining the context – Case study.

The considered organisation is BIOFOS, managing the wastewater treatment in the city of Copenhagen. One of the BIOFOS's main objective is the reduction of the pollution of the environment deriving from the treatment activities. Risk management is applied to deal with cyber-physical attacks which may damage the environment when the infrastructure of provided services are not sufficiently protected. The WWTP of the case study is characterized by four lines where all the biological treatment steps are run in parallel. The capacity of each line is equal to 2.500 m³/h, for a total treatment capacity of 10.000 m³/h. To cover the high peaks of inflow due to the rain events, the WWTP is equipped with equalization tanks with a volume of 44.000 m³. In Table A.3, the characteristics of the system under investigation are reported.

Table A.3. Characterization of the considered water system

Characteristics of the System	Values	Units
Number of Treatment Lines of the WWTP	4	[-]
Capacity of each Treatment Line	2.500	[m ³ /h]
Total Volume of the Equalization Tanks	44.000	[m ³]
Time to restore the Line under Maintenance	24	[h]

During and just after the rain events, given a certain hour of the day, when the actual inflow value is much higher than the expected wastewater flow, the inflow can be considered enough diluted, and a biological treatment could be by-passed without any significant environmental consequences on the receiving water body. This concept has been derived from common design criteria of CSO (Combined Sewer Overflow) devices in Europe. Specifically, a dilution coefficient r - given by the ratio between the actual total inflow (sum of rain Q_r and wastewater Q_{ww} contributions) and the wastewater flow (without any rain contributions) - with a value ranging from 3 to 6 is often considered as the minimum critical value r_c under which all the inflows should be properly treated.

$$r = (Q_{ww} + Q_r) / Q_{ww}$$

Obviously, in the phase of CSO devices design, the adoption of lower values of the critical dilution coefficient r_c leads to an increase of CSOs and a consequent decrease of treated volumes. In the considered case study, when the four lines are available and under operation, all undiluted wastewater is biological treated when considering a value of the critical dilution coefficient r_c equal to 3 even without the use of the equalization tanks, while all undiluted wastewater is biological treated when considering a value of the critical dilution coefficient r_c equal to 6 only by using the equalization tanks.

When the system has reduced capacity due to maintenance operations, the equalization tanks might not be enough to cover water overloads, since without usual capacity the plant might not be able to treat all the wastewater characterized by dilution coefficients below the designed r_c . Knowing in advance the expected inflow of the next 48 hours, the analysed DWC solution would help in planning the maintenance on each of the WWTP lines, without having consequences in terms of undiluted and untreated overflows of wastewater, despite the reduced treatment capacity available during the maintenance operations. Usually, the maintenance is performed during dry weather because in this case, in terms of capacity, there is not necessity of using all the four parallel treatment lines of the WWTP. The critical condition occurs if the attack is performed few hours before or during small-medium rain events. In this case the operators might perform maintenance expecting dry weather conditions, but they will eventually experience higher inflow than expected because of the considered attack, thus this will potentially lead to not treat wastewater not sufficiently diluted, according to the designed r_c .

The water organisation wants to identify the most hazardous scenario of pollutant concentration, where eventually, under cyber-attack, the polluted overflows released in the environment might have a dilution coefficient below 3.

On top of the Risk Management steps, at least one risk criterion must be defined, according to the objectives of the involved organization. For the case study, the yearly cubic meters (m^3 /year) related to biologically untreated and not diluted volumes (under the selected threshold of dilution coefficient) of wastewater were considered within a maximum duration of 24 hours per event.

The constrain of 24 hours per event was considered because for the specific identified risk, the organization stated that an eventual emergency can be recovered within 24 hours.

Note that each considered gate valve of the WWTP is being maintained once every two years and given that there are four treatment lines, the maintenance on one of the gate valves is being executed twice per year in average, according to the WWTP manager.

The value of the selected KPI is the yearly maximum of the mentioned polluted cubic meters related to the worst event with a maximum duration of 24 hours.

Based on the internal objectives of the organization, thresholds of levels of Risk have been defined, according to Table A.4

Table A.4. Pre-defined level of Risk expressed in terms of the selected KPI

Low Risk	Medium Risk	High Risk
KPI ≤ 120	120 < KPI ≤ 1.200	KPI > 1.200

The threshold between medium and high risk has been set to 1.200 m³ of undiluted wastewater, corresponding to the estimated minimum wastewater inflow entering the WWTP for one hour during dry weather. The threshold between low and medium risk has been set to 120 m³ of undiluted wastewater, corresponding to the estimated minimum wastewater inflow entering the WWTP for ten minutes during dry weather.

Hence, thresholds may trigger the Risk Treatment step, implying that the organization should consider risk mitigation measures to be implemented if thresholds are overpassed with the current measures already in place on physical and IT systems.

A.3. Risk Identification

Risk Identification methods within ISO framework. Since the objective of the risk management, as defined at step 1, is cyber-physical security of water systems, the risk identification step consists in the identification of physical, cyber, or physical-cyber risk events that will eventually have negative impacts on the achievement of the water utilities service goals. Before analyzing and evaluating the level of risk due to a certain cyber-physical attack in a water system, different aspects of that risk should be clarified. A proper risk description should comprise four elements, namely sources, events, causes, and consequences. Information sources useful to support risk identification include expert knowledge and judgement, personal and organizational experiences, checklists, historical records, incident databases, previous risk registers, and reports from previous risk assessments. Identifying the events and their possible paths is an important and not straightforward step in the Risk Management process. Each risk event has its causes and understanding them can significantly help in estimating the level of risk. The risk causes, type of threat and consequences are parts of each event path.

To facilitate the risk identification step, it is recommended to use a sheet or software that will list assets, threats, and vulnerabilities, as well as some other information like risk ID, risk owners, impact and likelihood, etc. To this purpose a Risk Identification Database (RIDB) has been developed in DWC to support water utilities in performing this step (DWC - D4.2, 2021).

The DWC RIDB identifies the type of threats, the sources of risk, the description of the events and the type of consequences produced. The RIDB covers the events identified by DWC partners as the most relevant risks related to their digital solutions developed in the project. The events included in the RIDB should be considered as individual "building blocks" from which the complex risk scenarios can be derived by their combination. Therefore, the RIDB does not include events generated by the combination of multiple risks. The RIDB of the DWC project builds on the approach provided by the STOP-IT project (STOP-IT – D3.2, 2018) which focused on cyber-physical attacks in water supply systems. The RIDB contains information on general possible risk events and may be used by water

organizations as a catalogue or collection of examples to identify their own specific risks. The RIDB is available at the following link: <https://risk-explorer.digital-water.city>.

To ensure consistency between the different events composing the RIDB, a specific sentence structure has been designed for the general description of the event:

A generates a B caused C of D affecting E, which might lead to a F issue.

Where:

- A: Type of risk source;
- B: Type of threat;
- C: Type of event;
- D: Specific asset;
- E: Type of asset;
- F: Consequences dimension.

The user can create a new event of the RIDB if the risk event of interest is not included in the RIDB, maintaining the same structure in each record.

Risk identification - Case study. General description of Copenhagen's digital solutions and identified risks.

As use case, a digital solution of the DWC project is considered (DS13 - Web platform for integrated sewer and wastewater treatment plant control). This solution allows a water utility to visualize the predictions of the flow entering in the WWTP with 48 hours in advance. Accurate rain forecasts are the input for a Machine Learning (ML) model which provides the timeseries of the mentioned flow, allowing optimized operations on the treatment process and improved schedules of maintenance. Since the water organization performs periodically programmatic maintenance on one of the four parallel lines of the WWTP, according to the DS's related partner, an internal attacker could manipulate the visualization of data and lead to a wrong flow prediction, thus to a wrong maintenance schedule. Specifically, the maintenance would be performed when dry time is expected. If a rain event is expected, the WWTP should be entirely working because of the expected high loads to be treated. However, the internal attacker may modify data to hide a rain event and the corresponding flow predictions. The water organization would start the programmed maintenance on one of gate valves of the lines.

Therefore, if this internal attacker knows about this planned maintenance and he/she manages to change the Web Interface visualization for the following 48 hours into a typical condition of dry weather, some troubles may arise because of the unexpected inflow, until the water organization will not restore full capacity recovering the line under maintenance within 24 hours.

Therefore, the WWTP with one line under maintenance is considered during expected dry time, and here is considered a relevant activity, e.g., maintenance on the inlet (or outlet) gate controlling the inflow (or outflow) in a treatment line which should be periodically maintained, so a quarter of the treatment capacity is missing. The time to restore the line under maintenance for regular operations can take up 24 hours, according to the water utility. The planned maintenance is normally done only

during dry weather, thus no effect of effluent quality under dry weather arises. However, under rain conditions, the hydraulic capacity of the biology treatment step is limited to 7.500 m³/h in this case. In the cyber-attack scenario, the actual rain has been hidden, thus an unexpected discharge overloads the three lines of the plant left in operation.

Exploring the DWC-RIDB, the water organization recognized a risk event generated by an internal attacker which could lead to quantity or quality issues on the effluent of the WWTP. Specifically, row 14 of RIDB is about the spoofing of the web application DS13 generated by an internal attacker. As reported in the database, considering the specific sentence structure related to each listed event, the risk is described as follows:

Internal attacker (A) generates a cyber threat (B) causing a Spoofing (C) of the Web Application of the DS 13 Interface (D) which affects Sewers or Wastewater treatment plant (E) and might lead to a Quantity issue (F).

Reminding that:

- A: Type of risk source;
- B: Type of threat;
- C: Type of event;
- D: Specific asset;
- E: Type of asset;
- F: Consequences dimension.

A.4. Risk Analysis

Risk Analysis methods within ISO framework. After having identified the risk events, the phase of Risk Analysis should be undertaken. During this step, the probabilities and/or the consequences are examined to assess the frequency and the impact on the water system of the identified risks. The Risk Analysis will also determine which risk factors (<https://www.merriam-webster.com/dictionary/risk%20factor>) would potentially have a greater impact on a water system. In this step, understanding how to model the risk event is key, keeping in mind which data would be required in the analysis in relation to the risk criteria set a-priori, and which variables are the most relevant for the identified risk event (e.g., potential critical areas, number of affected individuals, etc.). For instance, in the water infrastructure domain, a methodology which involves the stress-testing of drinking water supply systems have been developed in the STOP-IT project through the RAET (Risk Assessment and Evaluation Toolkit) (STOP-IT - D4.2, 2019). The stress-testing platform (STP) (Nikolopoulos et al., 2020) integrated in RAET can simulate both physical and cyber sub-systems coupling the simulation environment for the physical layer to an emulation environment able to model the cyber layer of the water system control and communication infrastructure (e.g., from SCADA to PLCs to monitoring), where cyber protection solutions will be implemented, and cyber-attacks attempted. The platform allows to analyse for example the effects of introducing malware to the supervisory system and trace these effects to Key Performance Indicators.

For risk analysis, there are three types of methods used for determining the level of risk, namely qualitative, semi-quantitative, and quantitative methods.

Qualitative Methods. This type of approach is often adopted for decision-making based mainly on expert judgment, experience, and intuition. These methods can be used when the level of risk is low and does not warrant the time and resources necessary for making an extensive analysis. These methods are also used when the numerical data available are not adequate for a more computational and quantitative analysis, so it would serve as the basis for a subsequent and more detailed analysis. The qualitative methods include brainstorming, questionnaires and interviews, evaluation for multidisciplinary groups, judgment of specialists and experts, etc.

Semi-Quantitative Methods. In this type of approach, classifications and scores are usually adopted, based on both estimations and computations on the ranges of likelihood and consequence of a certain risk event. These classifications are shown in relation to an appropriate scale for calculating the level of risk. High attention should be given with respect to the adopted scale, in order to avoid misunderstandings or misinterpretations of the results of the calculation.

Quantitative Methods. This type of approach allows to assign non-discrete values of loss to the various risks identified, enabling the calculation of the level of risk for several scenarios of attack. The quantitative methods include analysis of likelihood and consequences usually computed by multiple simulations. The quantitative assessment can be carried out through different approaches (e.g., Monte Carlo simulations, stress-testing combined with statistical analysis). The assessment of the consequences could be expressed in terms of KPIs related to different dimensions (finance, health, reputation, environment, etc.), depending on the nature of the risk that the utility is interested in (defined in the step of context definition). Given a digital twin of the system, stress-testing can be adopted as a method to compute potential impact given the cyber-attack resulted successful. On the other hand, if historical data are available, the probability of a successful cyber-attack could be computed based on the recognized past malicious events. Combining the objectively estimated probabilities with the consequences computed through stress testing simulations, a quantitative risk analysis can be performed.

The evaluation of the probabilities of successful cyber-attacks on water infrastructures could be more challenging when a new digital solution has been recently developed and/or no historical records about past events are available. To tackle this common issue related to the low level of awareness around emerging cyber threats, a semi-quantitative analysis can be performed. In this case, the impact can be still evaluated with a stress-testing approach, but probabilities can be evaluated based on a structured subjective assessment which should take into account expert judgments about multiple aspects such as the attractiveness of the assets or the hacking capabilities of the attacker with respect to vulnerable parts of the existing IT systems. Examples of semi-qualitative approaches for vulnerability of systems to cyber-physical attacks have been developed in the STOP-IT project with InfraRisk-CP (InfraRisk CP – User's guide, 2020). In InfraRisk, to assess the frequency of a successful attack to the water system the following approach is followed:

1. To find the frequency of an attack attempt (sometimes referred to as likelihood of threat happening) a set of questions is provided.
2. For each question there is a predefined list of answers, where each answer is associated with a score.
3. The scores are aggregated to give a total score for the frequency of an attempt.
4. To transform the score to a frequency number a low value f_L and a high value f_H are defined. f_L represents the frequency of an attack attempt if all scores for the attack attempt questions have the lowest possible values, and f_H represents the frequency of an attack attempt if all scores have the highest possible values.

5. To find the probability of the success of an attack attempt (sometimes referred to as likelihood of threat succeeding) another set of questions is provided.

6. For each of these questions there is also a predefined list of answers, where each answer is associated with a score.

7. To transform the score to a probability number a low value p_L and a high value p_H are defined. p_L represents the probability of a successful attack attempt if all scores have the lowest possible values, and p_H represents the probability if all scores have the highest possible values.

8. To find the frequency of a successful attack attempt, the frequency of an attack attempt is multiplied with the probability of success.

For assessing frequencies of cyber-attacks, a list of questions is provided, where scores are obtained for each sub question (S_1 , S_2 , etc.). If no information is available a score of 3 is given. If a question is not considered relevant, the score is excluded from the aggregation.

Risk Analysis - Case study. Copenhagen case study:

The analysis concerns with the quantification of the potential environmental impacts (consequence side) and the semi-quantitative estimation of the probabilities of successful cyber - attacks (probability side) leading to the discharge of potential yearly amount of undiluted, not stored, and biologically untreated water.

The impact has been assessed through the stress-testing of a digital twin of the system, and the probability of a successful attack has been evaluated through a structured subjective assessment since no historical records about malicious events are available.

- **Impact assessment**

Stress-testing of the system has been adopted as approach in the use case. A digital twin of the system has been adopted, especially taking into account the capacities of treatment lines and equalization tanks. Stress-testing consisted in exploring the response of the digital twin of the system under the maintenance of one treatment line. Specifically, all the available historical records of the inflow to the system have been considered as input and non-diluted and biologically untreated cubic meters of wastewater have been considered as output. In Figure A.2, the available inflows of 2020 at the inlet to the WWTP are shown.

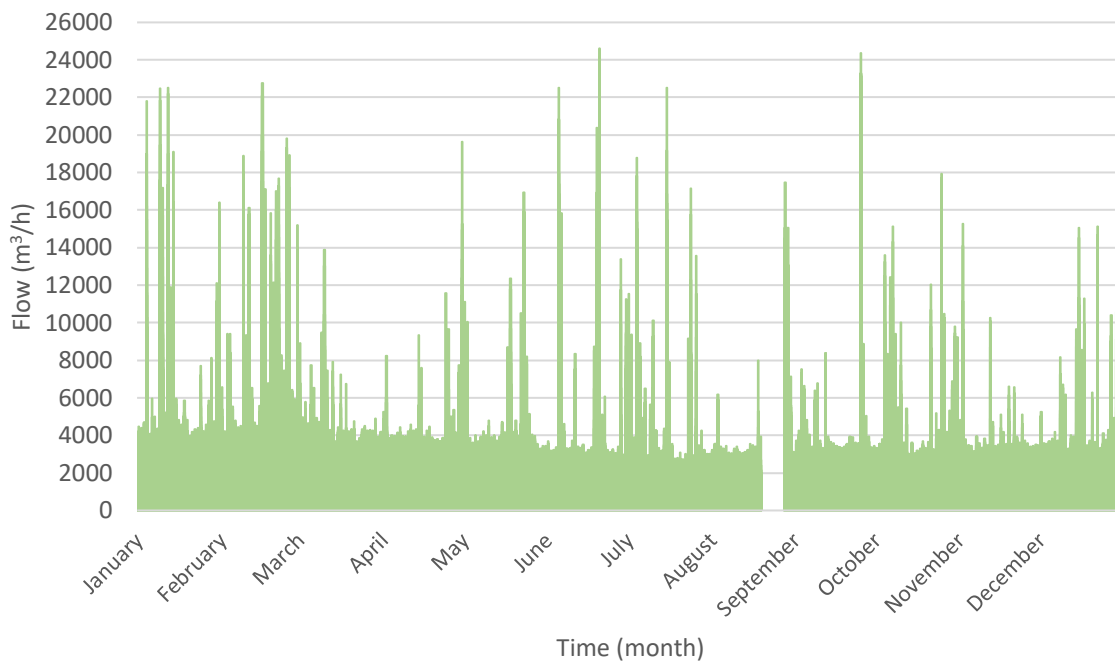


Figure A.2. Measured flows at the entrance of the analyzed WWTP in the year 2020

The available data, consisted in inflow values of the year 2020, with a resolution of one minute. The data shown in Figure A.2 were first aggregated with hourly averages.

Moreover, the missing data of part of the second half of August were excluded from the computations. According to the suggestions of the WWTP manager, one week of the first part of August (from 09.08.2020 to 16.08.2020) was considered as a reference for the weekly average wastewater flows without any rain flows contributions. In general, during other periods the inflow was higher because of frequent rain events and related high soil moisture levels which increased the run-off, arriving to the WWTP even with many hours of delay with respect to the time of actual rain precipitations. In Figure A.3, the hourly values obtained by averaging the corresponding flows of the considered week are depicted.

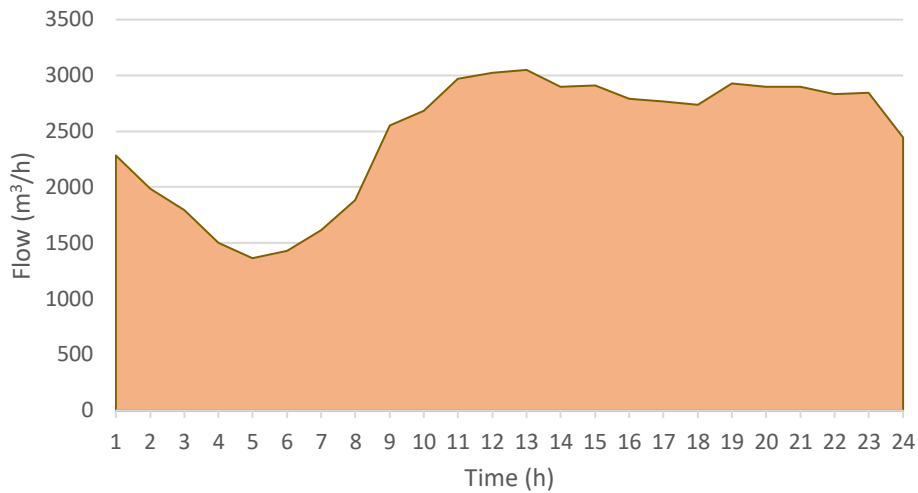


Figure A.3. Average daily flow pattern in dry weather conditions

Taking Figure A.3 as a reference for the entire 2020, dilution coefficients were computed along the year. On the basis of the data of the Figure A.2 and Figure A.3, the hourly dilution coefficients are shown in Figure A.4.

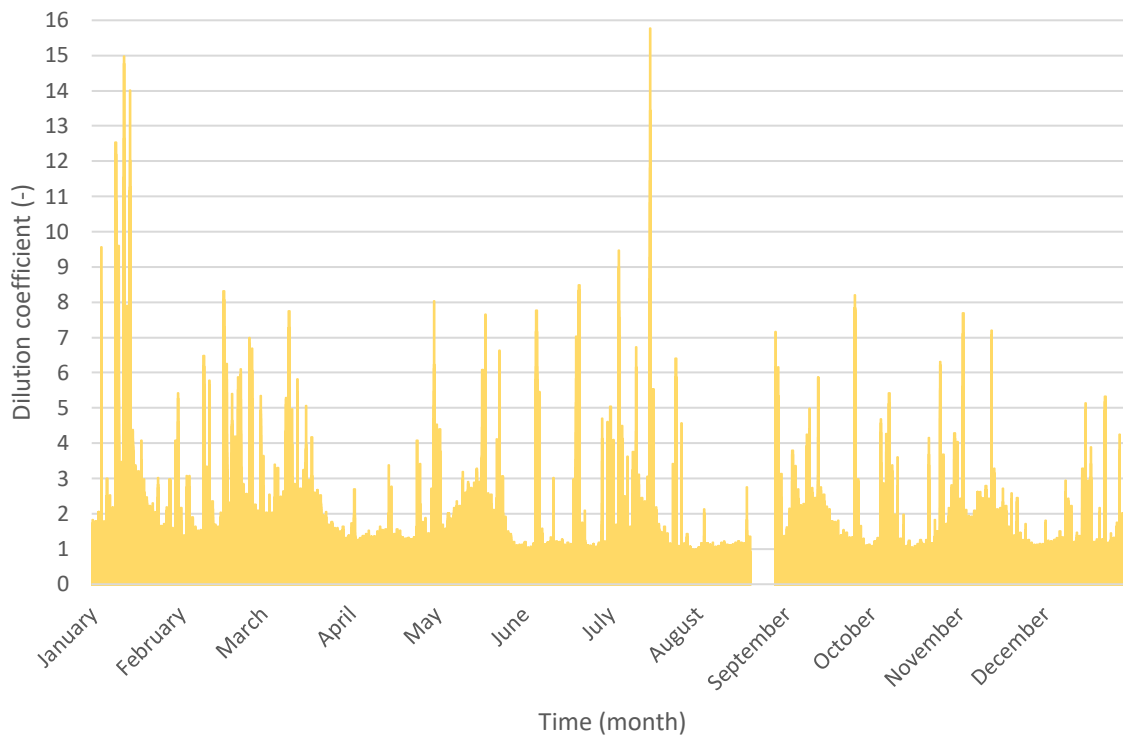


Figure A.4. Estimated dilution coefficients at the entrance of the analyzed WWTP in the year 2020

Concerning the critical dilution coefficient r_c , within the mentioned range of values between 3 and 6, the lowest value equal to 3 was considered, however the level of accepted dilution can vary from case

to case. Flows with a dilution coefficient equal or less than 3 corresponds usually to small rain events which could eventually leads to quantity and quality issues if part of the WWTP was wrongly scheduled for maintenance, so that the available treatment capacity is reduced. The 87 hourly values of the flow exceeding the capacity of 7500 m³/h, and which have a value of dilution coefficient r less than 3 are reported in Figure A.5.

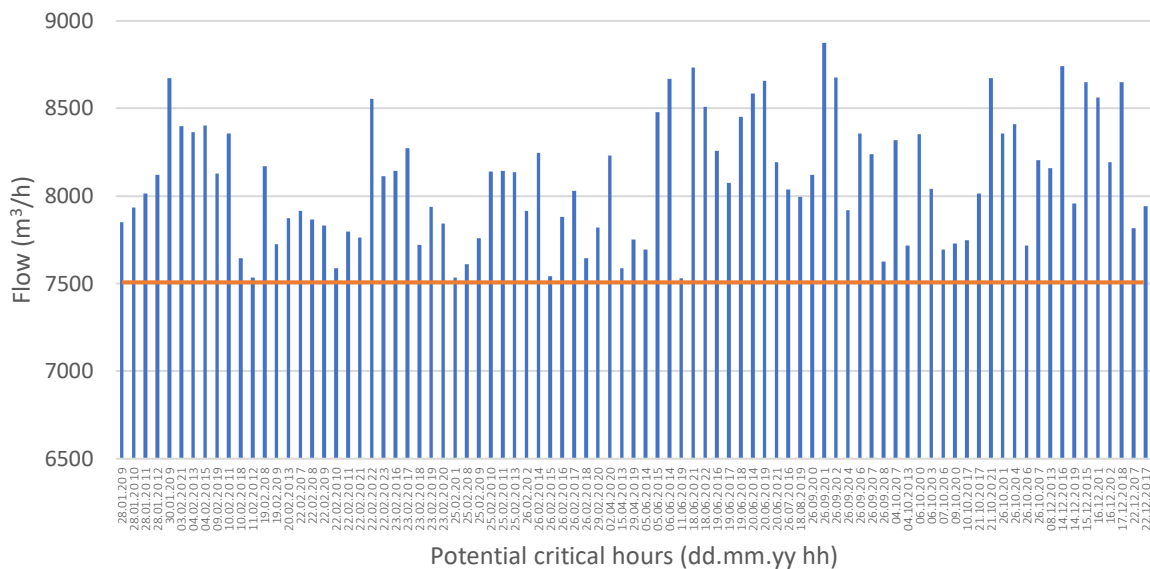


Figure A.5. Hourly flow values in 2020 greater than 7500 m³/h and coefficient dilution less than 3

The 87 hourly values showed in Figure A.5 could lead potentially to quality and quantity issues, but the equalization tanks can be used to store the wastewater if it is not completely full.

Looking at Figure A.5, the minimal accepted value of critical coefficient dilution equal to 3 leads to never overcome 10.000 m³/h, which means that under normal operations (without maintenance on one of the treatment lines) there are not events which can generate issues if the critical dilution coefficient is set to such value, nor the equalization tank must be used. On the other hand, if a higher critical dilution coefficient is considered (e.g., $r_c = 6$), potential issues arise even if all the capacity of the treatment lines is available, hence the equalization tanks are crucial also in normal operations.

With reduced capacity of 7.500 m³/h, even a critical dilution coefficient r_c equal to 3 leads to the mandatory use of the equalization tank. The tank was considered completely empty (thus with full capacity) on the 01.01.2020 and the water was considered as stored every time the flow at the entrance of the WWTP was greater than 7500 m³/h and released to the available treatment lines when the WWTP received less than 7500 m³/h. The available volumes in the process tanks of the WWTP were considered negligible with respect to the volume of the equalization tanks.

In the Figure A.6, the values of volume which should be stored, i.e., needed storage (NS), considering a treatment capacity equal to 7500 m³/h, are shown as blue bars and the maximum storage volume (SV) of the equalization tanks (equal to 44.000 m³) is highlighted with a horizontal orange line.

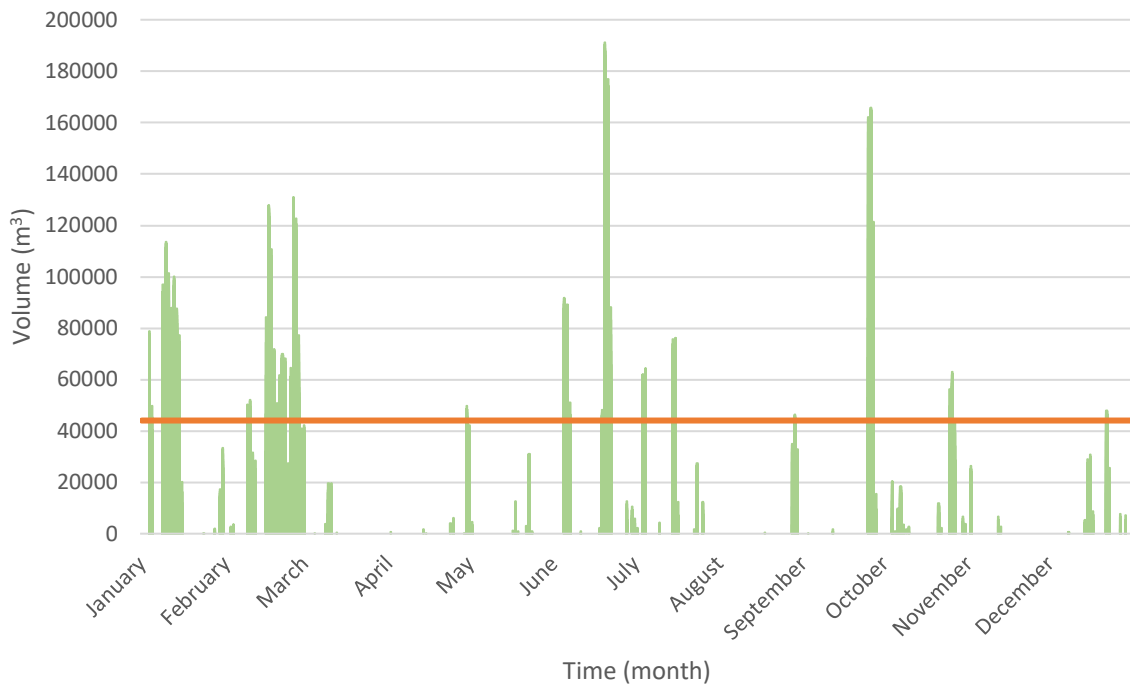


Figure A.6. Hourly values to be stored in 2020 with flow greater than 7500 m³/h and coefficient dilution less than 3

All the hourly values above the orange line of Figure A.6 could not be retained by the existing equalization tanks because they would overpass their capacity limits.

The dates of events shown in Figure A.5, which potentially could lead to quality and quantity issues, were matched with the events exceeding the maximum capacity of the equalization tanks shown in Figure A.6, as illustrated in Figure A.7.

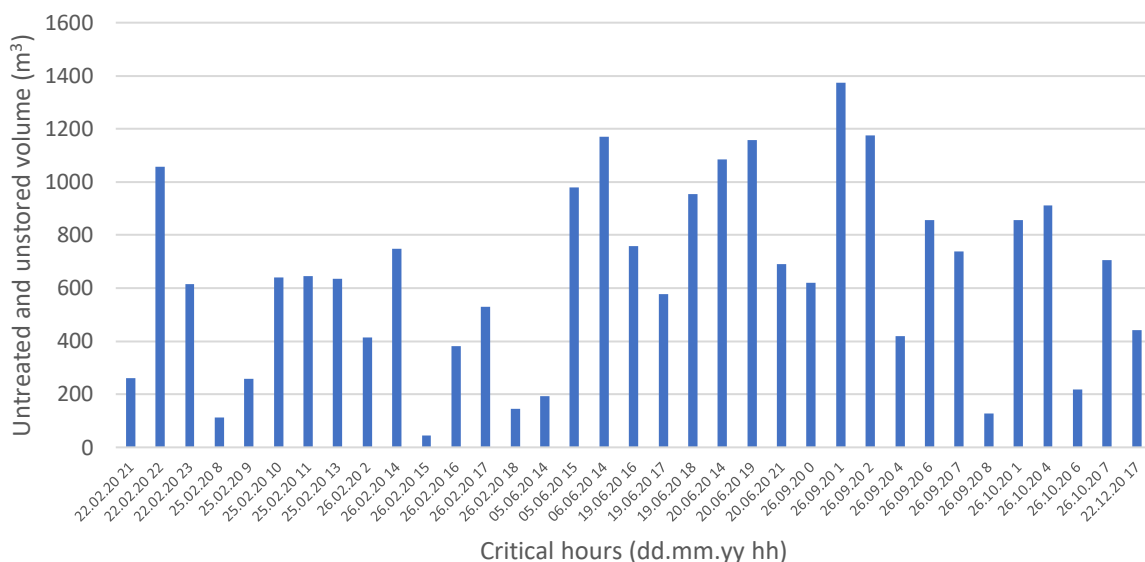


Figure A.7. Critical hourly values in 2020 not diluted, not stored, and not biologically treated.

The 35 hourly values shown in Figure A.7 (characterized by $r \leq 3$, Flow > 7500 m³/h, and NS > 44.000 m³), can be grouped in t events which last for less than 24 hours.

Under wrongly scheduled maintenance, all the events reported in Figure A.7 are related to significant issues because they are characterized by dilution coefficients equal to or below 3.

The selected KPI was evaluated. Note that, based on the dilution coefficients of the identified events, the untreated and not stored wastewater volumes were computed by dividing the volumes shown in Figure A.7 for the related hourly dilution coefficients reported in Figure A.4. In Figure A.8, the volumes of undiluted, untreated, and not stored wastewater are shown.

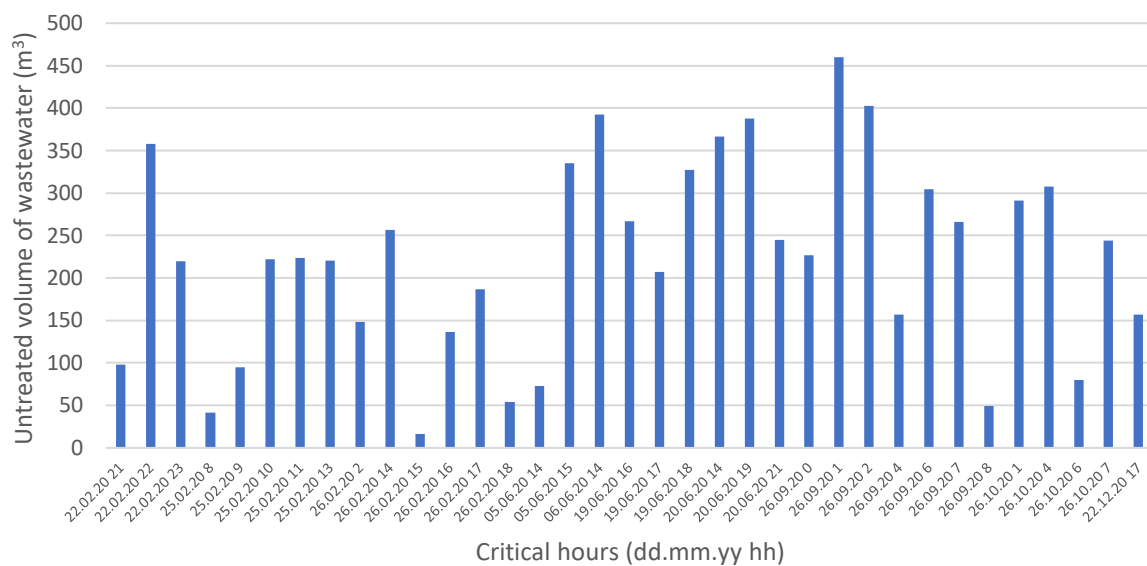


Figure A.8. Critical hourly values in 2020 of wastewater without any rain contribution.

Figure A.8 reports the events of 22nd, 25th-26th of February (more than 24 hours), 5th-6th and 19th-20th (more than 24 hours) of June, 26th of September, 26th of October, and 22nd of December.

The maximum value of untreated volume of the year within a maximum duration of 24 hours provides the value of the selected KPI, in terms of consequences. Specifically, among the mentioned 7 events, the 26th of September 2020 is associated with the maximum untreated volume of wastewater within the same 24 hours, equal to 1.865 m³ and which represents the consequence KPI value for 2020.

Looking at the dates of the critical flow events, a retrospective assessment was performed with respect to the rain events of 2020 to better understand the involved risk factors. In Figure A.9, the available rain events of 2020 for seven relevant stations (in the area of the studied catchment) of the Danish rain gauge network (SVK), expressed in µm/s and with a resolution of one minute, are reported.

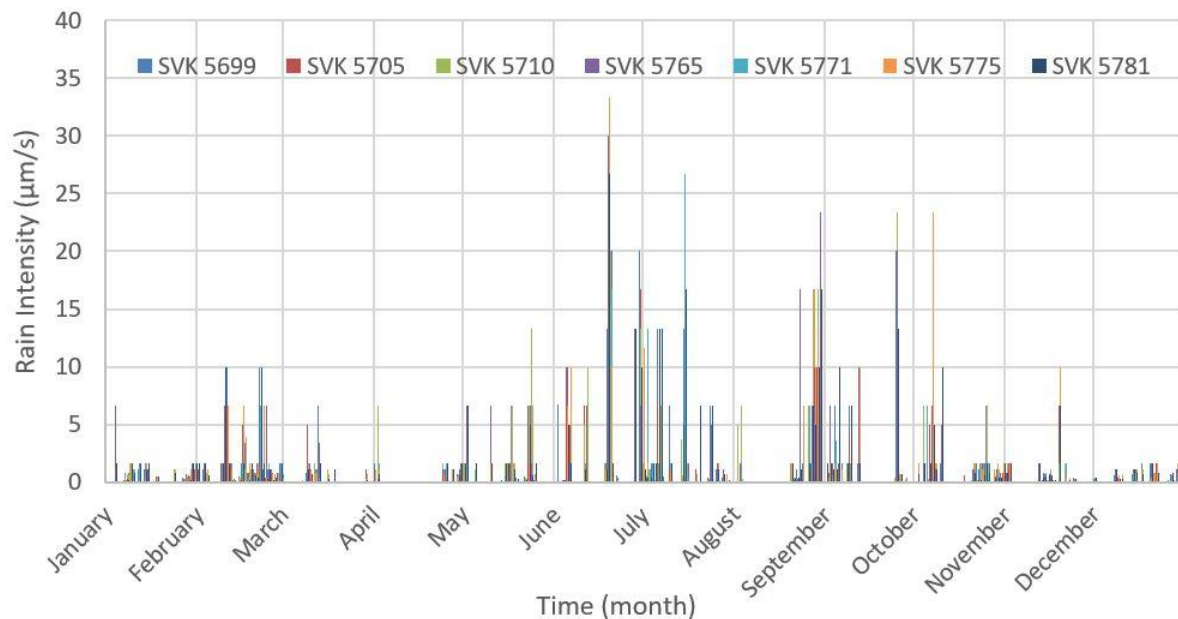


Figure A.9. Rain data in 2020 of seven station of the Danish rain gauge network.

In terms of risk factors, it is important to highlight that the most dangerous conditions are not necessarily connected to the most extreme rain events (see for instance 22nd, 25th-26th of February, 5th-6th of June, 26th of October, and 22nd of December) mainly because the values of dilution coefficients would be higher than 3 for a large part of the event, so wastewater could be considered enough diluted, and the lack of biological treatment would not be a significant environmental issue; however, on the other hand, the equalization tanks might be more easily filled completely during extreme events.

Moreover, extreme rains are more likely detected in advance through additional sources of information, thus for the attacker it would be more difficult to fool the WWTP operators and plan the attack during a scheduled maintenance just before a well-known and expected extreme event.

- Probability assessment

To evaluate probabilities or the frequency of a successful attack, the approach for cyber-attacks proposed in InfraRisk-CP from the STOP-IT project was considered. The document (both in .csv and .xlsx extension) at the link <https://zenodo.org/record/6589610#.YpIB4aBBw2w> supports calculations of InfraRisk-CP to estimate the probability of a successful cyber-attack, as reported in this sub-chapter. In the Annex, the details of InfraRisk's approach related to physical attacks is reported. It was assumed that the internal attacker is aware of the most favorable conditions for an attack, in terms of expected flows and planned maintenance in the system.

Each considered gate valve of the WWTP is being maintained once every two years and given that there are four treatment lines, the maintenance on one of the gate valves is being executed twice per year in average, according to the WWTP manager.

Since the attacker is internal, it is assumed he/she has the possibility to drive **partially** the schedules of the programmed maintenance at the same moment of the 7 events during the year which produce consequences for the analysed risk. The term "partially" is due to the extent of the attacker's will to drive the maintenance schedules and carry out the attack, which depends on the scores (S1-S15) of the InfraRisk methodology. Thus, regardless of the attack attractiveness, assuming that over a period of 10 years the attacker would try at least one attempt, his/her frequency of attack spans from a minimum of once per ten year and a maximum of two times per years.

The frequency between these two extremes values for attempting the Spoofing of the Web Application of the considered Digital Solution Interface can be estimated through InfraRisk.

- fL (Lowest attack frequency) = 0.1/year
- fH (Highest attack frequency) = 2/year

The probability of success is mainly related to the capabilities of the attacker and to the security of the IT system. According to BIOFOS, the attacker is supposed to have good chances to penetrate the IT system since the considered attack is internal, but the actual value mainly depends on the permissions already owned by the attacker and his/her capacities. Specifically, if the attacker already has the full permission to the company's IT system, as an *administrative user* and the IT system has negligible protection in comparison with his/her capabilities, the probability is estimated to 100%. On the other hand, if the attacker is a *technical user* of the company and his/her capabilities are negligible in comparison with the IT system, some effort on stealing the required credential of an *administrative user* would be needed, thus in this case the probability is estimated equal to 1%.

- pL (Lowest attack success probability) = 1%
- pH (Highest attack success probability) = 100%

In the following, the answers of BIOFOS to InfraRisk-CP questions are provided.

1) How attractive it is to make an attempt to attack the water system, in terms of:

- Recognisability?

Answer: **S1 = 2 (low)**, due to the fact that there is no recognisability in affecting the wastewater treatment plants, power plants and distribution system are at a much higher risk.

- Symbolism?

Answer: **S2 = 1 (very low)**, this will likely not affect the citizens, but 'only' the environment, drinking water and distribution systems are at a much higher risk the wastewater treatment plants.

- Potential for economic profit (e.g., ransom)?

Answer: **S3 = 3 (medium)**, Organized crime does not specifically target wastewater treatment plants, but there is a medium risk.

- **Potential for political profit?**

Answer: **S4 = 1 (very low)**, Other utility sectors are at a much higher risk, electricity/power and drinking water utilities.

Note: *Recognisability* deals with attackers having a desired to be recognized within a community. Typically, this could be individual hackers. *Symbolism* could be relevant for terrorist groups which often have an objective to cause fear and uncertainty. Economic profit would relate to organized crime. Political issues could relate to foreign nations or political groups within one nation.

The scores S1, S2, S3, and S4 could be seen as competing scores, and we let be a total attractiveness score $S_A = \max(S1, S2, S3, S4) + \Delta_A$. Here, $\Delta_A = 0.25 \ln n$, where n counts the number of scores equal the maximum score. Note that $\Delta_A = 0$ if the maximum score is 1 or 5 or the maximum score appears only once. In the analysed case **S_A is equal to 3**.

2) Level of Organizational issues, specifically regarding:

- **Measures implemented towards insiders?**

Answer: **S5 = 4 (low)**, scarce employees' education regarding implemented IT security. User accounts for system access are in place, but no internal system to catch unsuccessful login/or hacking attempts.

- **Quality of internal surveillance and intelligence systems?**

Answer: **S6 = 4 (low)**, no central system is implemented.

- **Systematic preparedness exercises, investigation, and learning?**

Answer: **S7 = 5 (very low)**, never completed an exercise on the IT systems and infrastructure.

- **Security focus in agreements with vendors and contractors?**

Answer: **S8 = 4 (low)**, vendors and contractors use to sign a confidentiality agreement regarding GDPR and information obtained during work/interaction with BIOFOS.

Note: For the organizational factors affecting the frequency of attack we calculate an average score: $S_o = (S5 + S6 + S7 + S8)/4$, so in the analysed case **S_o is equal to 4.25**.

3) Conditions affecting if an attacker will make an attack attempt for a specific component:

- **How vulnerable the component seems from the attacker's point of view?**

Answer: **S9 = 2 (low)**, technical systems are behind the company firewall and a technical firewall that covers all the technical IT-systems. No administrative IT system user has direct access to the technical systems. A different technical username is required.

- **Visible protective measures by the utility manager for the specific component.**

Answer: **S10 = 2 (high)**, low, physical access to buildings and components is restricted. Alarm systems in buildings.

- **How critical the component seems from the attacker's point of view?**

Answer: **S11 = 2 (low)**, normal attackers does not have specific knowledge regarding the operations, equipment and control used at the wastewater treatment plant

- **Accessibility of the particular component.**

Answer: **S12 = 2 (low)**, all technical computer terminals are locked when not in use. Components (motors, gates) at the treatment plant cannot be operated locally when in automatic control mode.

- **Attacker's capability vs required capability to make an attempt.**

Answer: **S13 = 3 (medium)**, an attacker needs some skills to make an attempt, but it is possible.

- **Attacker's available resources vs required resources.**

Answer: **S14 = 3 (medium)**, an attacker needs good resources to make an attempt, but it is possible.

Note: For the conditions influencing willingness of an attacker to make an attempt an average score is also proposed: $S_w = (S_9 + S_{10} + S_{11} + S_{12} + S_{13} + S_{14})/6$, so in the analysed case **S_w is equal to 2.33**.

4) Evidence with respect to possible attacks:

- **How is the actual cyber security situation evaluated by the security authorities (police, intelligence, etc.)?**

Answer: **S15 = 3 (medium)**, wastewater treatment plants are not the first in line for an attack, higher risk at power plants and power distribution and drinking water production and distribution.

- **Evidence from internal surveillance for the specific attack (computerized monitoring tools).**

This quantity is measured in terms of number of attack attempts per time unit, typically per year. Answer: **S16 not available**, main users cannot be currently detected, normal users would use workstation which are recognized; however, at the current time there are no evidence.

Note: To obtain a total normalized score for the likelihood of an attack, consider the average of S_A , S_O , S_w and S_{15} and standardize between 0 and 1:

$L = (S_A + S_O + S_w + S_{15} - 4) / (20-4)$, so in the analysed case **L is equal to 0.54**.

The frequency of an attack based on the influencing conditions is given by:

$$f = f_L \left(\frac{f_H}{f_L} \right)^L$$

The yearly frequency based on the assessment of conditions can be averaged with the observed frequency S_{16} , if available. In the analysed case **f is equal to 0.5/year**.

For the probability assessment of a successful attack another set of questions are provided.

5) Likelihood of succeeding in an attempt:

- **Attacker's capability vs required capability to succeed in an attempt**

Answer: **S17 = 4 (high)**, since the attacker is internal, but normally an attacker must overcome several firewalls and login to specific systems to succeed.

- **Attacker’s available resources vs required resources to succeed in an attempt**

Answer: **S18 = 4 (high)**, since the attacker is internal, but normally only highly trained attackers can access and penetrate the implemented security measures to gain access to technical systems.

- **Explicit protective measures**

Answer: **S19 = 2 (high)**, even if the attacker is internal because when using VPN access, encryption is used. Moreover, only VPN access from Danish IP addresses is allowed, a 2-step user verification for VPN access is adopted, and administrative IT user must login to VPN. To access the technical systems a technical user is allowed only via a VMware remote desktop, no direct server access. Finally, there are regular software updates of firewall, antivirus tools, clients, servers for both administrative and technical systems.

To obtain a probability measure for success of the attack, a standardised score is calculated in the interval from 0 to 1, with $Q = (S17 + S18 + S19 + S6+ S7 - 5)/20$, so in the analysed case **Q is equal to 0.7**.

Note that in this score two of the organizational conditions are included. It could be argued that this gives “double counting”, but after the normalization, this is considered not to be an issue. The probability of a successful attack is given by:

$$p = p_L \left(\frac{p_H}{p_L} \right)^Q$$

In the analysed case **p is equal to 0.25**.

The frequency of a successful attack is given by:

$$f_A = f \times p$$

This frequency is expressed as a certain value per year, so in the analysed case **f_A is equal to 0.125/year**, slightly more than once per ten years (Occasional).

In some situation it is desirable to assign the frequency of successful attack to a likelihood *category*. In InfraRisk, the following likelihood categories are defined:

- | | |
|-------------------|------------------------------|
| 1. Very unlikely: | Less than once per 100 years |
| 2. Remote: | Once per 10-100 years |
| 3. Occasional: | Once per 1-10 years |
| 4. Probable: | 1 to 12 times a year |
| 5. Frequent: | More than once a month |

These categories can be used to transform the estimated frequency to a category number when a risk matrix is adopted for the Risk Evaluation phase.

A.5. Risk Evaluation

Risk Evaluation methods within ISO framework. The Risk Evaluation step involves the comparison of the results of Risk Analysis with the risk criteria and KPIs target values established at step 1 to

determine the level of severity of the analysed risk events and therefore to make decisions on the need of taking remediation or prevention actions. This phase leads to a set of possible decisions, listed in the following:

- do nothing further;
- consider risk treatment options;
- undertake further analysis to better understand the risk;
- maintain existing controls;
- reconsider objectives.

Decisions should take into account the wider context and the actual and perceived consequences to external and internal stakeholders. The outcome of the Risk Evaluation should be recorded, communicated and then validated at appropriate levels of the organization.

To facilitate the evaluation, the response of the system to the assessed inputs should be related to risk criteria, selected on the basis of the water utilities goals with respect to the identified risk events. When risk reduction measures are implemented, the digital twin of the system should change accordingly, thus the step of risk analysis and associated risk evaluation should be run again.

After stress-testing the water system under different configurations of the identified risks and based on the selected KPIs, it is possible to derive the conditions which may lead to the most serious consequences on the infrastructure. After having combined the estimated consequences with their probability of occurrence, the comparison between the result of "Risk Analysis" phase and the target values set in "Defining the context" phase and the assessment of the eventual defined levels of risk are performed at this stage. The comparison and eventual definition of the level of risk can lead to the optimal selection of the risk reduction measures to be implemented within the Risk Treatment phase.

Risk Evaluation - Case study

The risk evaluation for the case study of Copenhagen was derived from the previous phase of risk analysis. Specifically, the worst event of 2020 which may cause environmental issues was considered for the consequence assessment, in terms of the selected KPI, equal to 1.865 m³/year.

The probability of a successful attack per year is equal to 0.125, and it was computed through the approach which follows the InfraRisk's methodology. Multiplying the consequences expressed in terms of the KPI with the probability, the risk is finally evaluated as Medium, by comparing the results with level of risk defined in Table A.4, since the risk is associated to the risk actual value (KPI) equal to 233 m³/year.

Table A.5. Identification of the level of risk by comparing results with targets values

Low Risk	Medium Risk	High Risk
KPI ≤ 120	120 < <u>KPI</u> ≤ 1.200	KPI > 1.200
No Low Risk: 233 > 120	Medium Risk: 120 < 233 ≤ 1.200	No High Risk: 233 < 1.200

Based on this evaluation, different risk reduction measures could be adopted in the next phase of risk management, consistently to the actual level of risk.

A.6. Risk Treatment

Risk Treatment methods within ISO framework. The purpose of risk treatment is to modify the risks derived during the risk assessment; in general terms, this can be performed by reducing the likelihood of an incident and/or by reducing the impact on the system. The Risk Treatment phase itself involves an iterative process made of the following steps:

- formulating and selecting risk treatment options;
- planning and implementing risk treatment;
- assessing the effectiveness of that treatment;
- deciding whether the remaining risk is acceptable;
- if not acceptable, taking further treatment.

Since the purpose of the Risk Treatment is to select and implement the best options for addressing the identified risk, the DWC project provides a Risk Reduction Measures Database (RRMD) (DWC - D4.2, 2021), where several risk reduction measures are collected similarly to the Risk Identification Database (RIDB), previously discussed.

The RRMD is a catalogue to assist risk managers in the process of finding suitable measures for an appropriate Risk Treatment which was firstly developed under the project STOP-IT (STOP-IT - D4.3, 2019). The aim of the collected measures is the reduction of existing risks that have been identified by the application of the RIDB. Since the RRMD is not and cannot be an exhaustive list of all possible RRM, it shall not supply a fully prepared and formulated plan for Risk Treatment, but rather show to the user options on how existing risks could be treated by choosing and implementing one or several measures. Thus, it is important to enable future users of the tool to populate the database with additional measures. Only by keeping the RRMD a “living register” its practical value can be ensured also in the future, also with respect to incoming cyber-physical threats of critical infrastructures, so the users may contribute by adding new relevant measures in the database. The DWC database has been populated with generally described Risk Reduction Measures (RRM) to extend the applicability to other water systems. This ensures the implementation of the listed measures in a large variety of cases that might differ from the external and internal contexts of the organizations involved in the DWC project. The RRMD is implemented similarly to the RIDB (<https://risk-explorer.digital-water.city>).

Many-to-many relationships between risk events of RIDB and measures of RRMD can be realized. Thus, an event of the RIDB may be associated to several suitable measures of the RRMD. On the other hand, a measure from the RRMD may address several risks documented in the RIDB.

When risk reduction measures are selected, the steps of risk analysis and risk evaluation should be performed by considering the change introduced by the selected measures. The benefit in terms of risk reduction can be compared to the cost of the considered measure, thus a final decision is taken.

Risk Treatment - Case study

By exploring the RRMD the following RRM have been selected by the user as potentially relevant to their case:

- Implementation of IT security systems
- Implementation of training procedure of the employees
- Increase of the volume of the equalization tank

In the first two cases, the risk is prevented, while in the third case the consequences are mitigated. The suggested complete list in the RRMD connected to the selected risk event is shown in Figure A.10

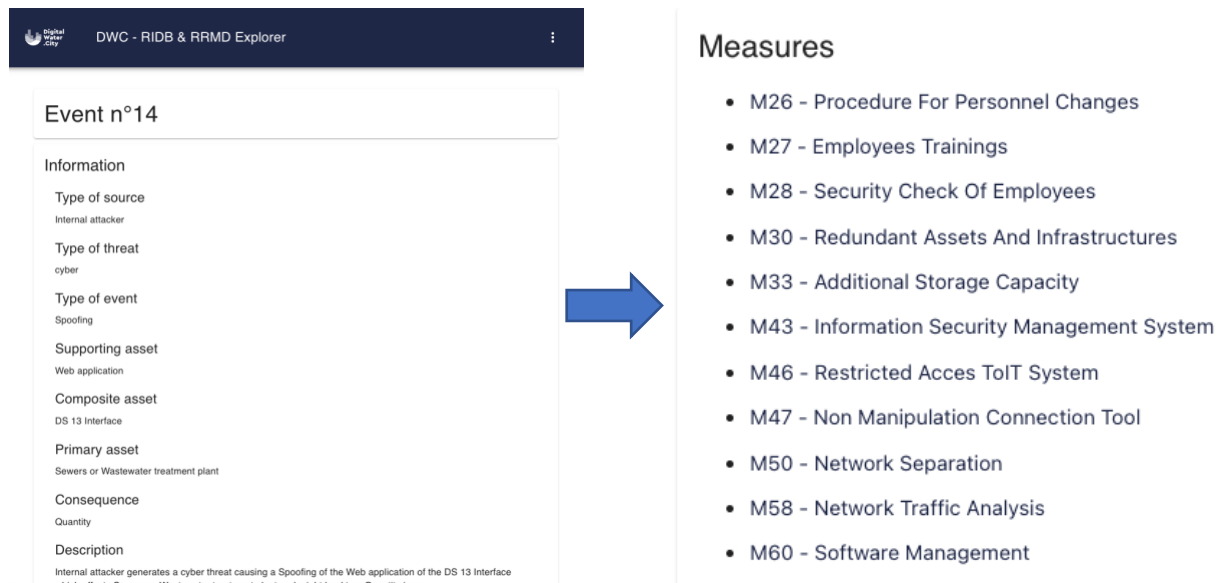


Figure A.10. Risk reduction measures in the RRMD associated with the identified risk in the RIDB.

The decreased probabilities of a successful attack when the first two types of measures are implemented are site-specific and depends on the level of the protection of the IT system.

In InfraRisk, if all the quantities which depend on the level of protection of the organization (S5-S6-S7-S8-S17-S18-S19) are raised to their best score, the obtained estimation of the probability of a successful attack is equal to 0.017, leading to significant reduction of the risk, since it would be decreased almost to the 14% of the original value of risk, i.e., 33 m³, correspondent to **Low Risk**. On the other hand, if the organization implements actions with decreased consequences (e.g., obtained with M33, for instance by doubling the volume of the equalization tanks), KPI could be computed through the same procedure of stress testing, described in the risk analysis part. By adopting a doubled storage volume in the equalization tank (M33, with 88.000 m³ of storage) and by following the same procedure described for the consequence assessment in the Risk Analysis paragraph, the hypothetical untreated polluted overflows of 2020 were computed and are reported in Figure A.11.

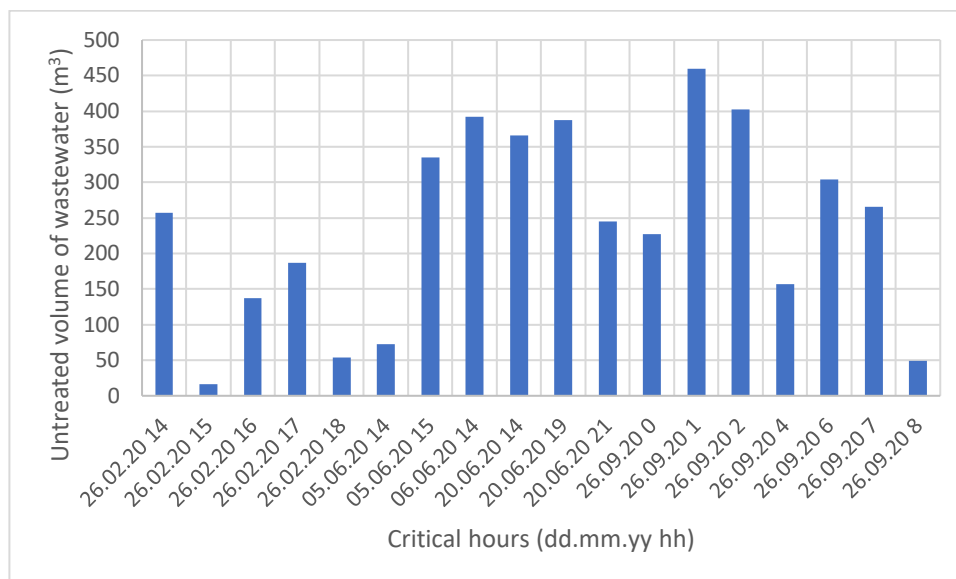


Figure A.11. Critical hourly values in 2020 of wastewater without any rain contribution in the case of doubling the volume of the equalization tanks as risk reduction measure.

Considering the definition of the selected KPI, its new value would result exactly the same as before the implementation of the equalization tank, equal to 1.865 m³ in terms of consequences and 233 m³ in terms of risk, showing how important is to take decisions on the basis of a risk analysis. In fact, even if the yearly untreated volume is globally reduced, the yearly maximum event, relevant to the performed risk assessment, does not report any significant improvement connected to this mitigation measure. On the contrary, if an additional treatment line with a capacity of 2.500 m³/h is installed (M30), no consequences would be reported, since all the yearly inflow can be treated during the eventual considered attack. In this case, the water organisation can rise its level of standards, for instance by increasing the minimum level of accepted dilution coefficient. In particular, a value of risk similar to the one computed with the actual conditions would be reached by considering a maximum acceptable dilution coefficient of 5 ca., instead of 3.

Moreover, the implementation of additional [IT security solutions](#), such as a proper firewall for the web application, would impact for instance scores S9, S12, and S13 of the InfraRisk assessment, providing less probability of successful attacks.

The decision about the risk reduction measures to be implemented is dependent on a cost-benefit analysis, thus is highly site-specific.

Nevertheless, because of the extremely high potential for risk reduction related to due to the reduction of the probability component, in comparison with the expected implementation cost, the following risk reduction measure is suggested for the case study:

- **Training procedure to reduce human errors**, in fact while not being an attack per se, human errors can lead to the same consequences. If a user is given access to data or actions, he/she should not have access to, he/she could misuse it (intentionally or not) and effectively create a situation similar to an attack.

By assuming an effective implementation of the mentioned measure to reduce probabilities of a successful attack, the score **S7** would pass from 5 (current situation) to 2 (e.g., exercises performed by all the employees on the protection of the infrastructure and IT system), leading to an estimated f_A

equal to 0.055/year, corresponding to a risk equal to 103 m³/year is estimated, i.e. **Low Risk**, considering the selected risk criteria.

A.7. ANNEX - InfraRisk-CP assessment for physical attacks (from the user guide of STOP-IT)

Frequency of physical attack

For physical attacks the following questions with possible answers are provided for the frequency of an attack:

Q1: How attractive is the asset to the perpetrator?

- 1=Very low attractiveness
- 2=Low attractiveness
- 3=Medium attractiveness
- 4=High attractiveness
- 5=Very high attractiveness

The attractiveness is influenced by the possible damage potential, the political, social and economic importance, the psychological effects as well as by the affected end-users (military institutions, parliaments, chemical industry, residence of important people like a president or similar). The classification of the attractiveness is done subjectively as for example the perpetrator is not known at the time the assessment is done.

Q2: How is the actual security situation evaluated by the security authorities?

- 1=Police intelligence does not expect any threats
- 3=Evidences for a threat exist
- 4=The asset is endangered, an attack cannot be excluded
- 5=The asset is in significant danger, an attack should be expected

Information about the actual security situation can usually be gained at the responsible police authority.

Q3: How relevant is the asset for the overall water supply?

- 1=Low
- 2=Medium
- 3=High
- 4=Very high
- 5=Critical

A systematic can be applied to answer this question. For example, each answer can be matched to a certain percentage of end-users/people (e.g., Low -> <10 %, Medium -> <25 %, ...). Other possibilities could be the matching of answers with percentages of the overall drinking water amount affected or similar. It might often be true that the asset is e.g., very relevant for a certain part of the network but only medium relevant for the overall network. In these cases, the relevance of the asset should be rated with regard to its importance for the affected part of the network.

Q4: How difficult is it to carry out a criminal act?

- 1=Extremely high effort necessary (1 point)
- 2=Medium effort necessary (2 points)

3=Little to no effort necessary (3 points)

The choice of an answer is based on the assumed effort of the perpetrator and the possibility to be successful with that assumed effort. For the evaluation the attack path of “lowest resistance” should be considered.

Q5: Do special environmental conditions exist that temporarily increase the need for protection?

1=No special conditions (1 point)

2=Few special conditions (2 points)

3=Substantial special conditions (3 points)

Here special temporarily occurring events or conditions are regarded. Examples could be government visits, major events, festivals, etc.

Now let $L1$ be the sum of scores achieved for questions Q1 to Q5. $L1$ can take values in the range 5 to 21, and a standardized score between 0 and 1 is given by: $L = (L1-5)/(21-5) = (L1-5)/16$. The frequency of a physical attack is now calculated similarly to the cyber-attack case.

Probability of physical attack succeeding

The following questions with possible answers are provided for the probability of a successful attack:

Q6: How is the asset built? Is it easily visible for the public?

3=Object not visible for public

2=Object visible without restrictions

1=Object only accessible by interruptions of the public life (railway, streets, ...)

For example, if an asset is built in a very enlivened area of a city an attack is more likely to be detected by people and thus, not succeeding compared to an attack on an asset that is built in a forest where a perpetrator is more or less undisturbed.

Q7: In which resistance class is the perimeter protection built?

2=RC1-RC2

1=RC3-RC4

0=RC5-RC6

The different resistance classes used in the possible answers are defined in DIN EN 1627 (DIN 2011). If there is any gap or similar in the perimeter protection, the score of 2 is given.

Q8: In which resistance classes are the walls of the buildings including their integrated integrations?

2=RC1-RC2

1=RC3-RC4

0=RC5-RC6

The different resistance classes used in the possible answers are defined in DIN EN 1627 (DIN 2011). If there is any gap or similar in the protection like unprotected windows lower than the first floor, the score of 2 is given.

Q9: How is the sensory surveillance realized?

2=No sensory surveillance

1=Binary Contacts, e.g., open/closed

0=Measured value-based surveillance, e.g., sensitivity of sensor can be regulated

The evaluation of the sensory surveillance should be realized at the weakest position of the barriers.

Q10: How are organizational measures implemented?

2=No organizational measures exist

1=Primary dissuasive measures are implemented like alarms, only irregular patrolling

0=Organizational measures ensure, that a direct defensive reaction is initiated (e.g., the police is called, the system is shut down, ...).

Now let $Q1$ denote the sum of scores for questions Q6-Q10. $Q1$ can take values in the range 1 to 11, and a standardized score between 0 and 1 is given by: $Q = (Q1-1) / (11-1)$. The probability of success of an attempt and the frequency of a successful attack are thus calculated similarly to the cyber-attack case.

Appendix B

IoT Security Checklist

The IoT Security checklist is a questionnaire-like document to be used for a self or guided assessment of an IoT device. The objective is to raise awareness on specific weaknesses. It aims to be domain agnostic. The questions come from both our experience working with IoT devices and guidelines such as the “Baseline Security Recommendations for Internet of Things in the context of critical information infrastructures” from ENISA [27].

Following the example of the [OWASP Application Security Verification Standard](#), three levels are defined:

■ **Level 1:** Level 1 is the bare minimum security IoT devices should strive for. Complying with this level should counter attackers who are using simple and low effort techniques to identify easy-to-find and easy-to-exploit vulnerabilities. In the case the IoT device is processing sensitive data or critical for operation, you probably don’t want to stop at this level.

■ **Level 2:** Level 2 aims to defend against the most common risks associated with IoT devices today. It is appropriate for devices processing healthcare data or other sensitive assets. Threats to level 2 are typically skilled and motivated attackers focusing on specific tools and techniques that are effective to discover and exploit weaknesses within application. Aiming at this level should be enough for most devices.

■ **Level 3:** Level 3 is typically reserved for devices requiring a significant level of security verification, such as in the military, health and safety or critical infrastructure domains. If you think your device must comply with level 3, then this checklist won’t be enough in itself (it can provide a good start though) and you probably want to also look at IoT certification schemes such as Common Criteria, FIPS-140 or PSA Certified.

Hardware

■ **Q1. Are the debug ports disabled?**

Debug ports often used in development such as JTAG or SWD can be used by an attacker with physical access to the device to acquire the firmware of the device (and thus potential cryptographic material) and perform reverse engineering of it to identify software vulnerabilities.

■ **Q2. Are interfaces allowing memory access disabled?**

Similar to Q1, interfaces allowing memory access such as DMA should be disabled if not required to prevent an attacker to access memory. DMA being a useful feature for many IoT applications, an alternative is to restrict the memory accessible to it, by configuring the Memory Protection Unit (MPU) for instance.

■ **Q3. Do you use tamper proofing mechanisms?**

Similarly to obfuscation (Q7), one should rely on tamperproofing for security as any device or system can be broken by a person with sufficient knowledge, time and resources. Having tamper proofing mechanisms can however discourage or slow down certain attackers. Tamper proofing should not rely on network connectivity.

■ **Q4. Is the device designed to make it difficult to access pins and electrical traces?**

An attacker can easily probe a microcontroller's pins if they are exposed (through TSOP chip package for instance). Prefer BGA package when possible and avoid exposing electrical traces on the PCB.

■ **Q5. Are you using any specialized security chips or coprocessors?**

Security chips / coprocessors integrate security at the chip level, providing trusted storage of device identity and authentication means. They also provide protection of device keys at rest and in use, and prevent unprivileged code from accessing security sensitive code [27]. Without this, an attacker will be able to retrieve device keys, potentially allowing him to impersonate the device, steal data and inject fake data into the system.

■ **Q6. Are you employing hardware-based immutable Root of Trust (RoT)?**

Without a RoT, there is nothing to prevent an attacker having physical access to the device to take over it, introduce unauthorized code, steal data or even to form botnets.

■ **Q7. Do you obfuscate the device's components?**

While it is not recommended to use obfuscation as a sole security mechanism (and to rely on it), using obfuscation can make the life of an attacker more difficult and potentially discourage a certain class of attackers (such as "script kiddies"). Obfuscation can consist of erasing/masking IC names, not labelling the PCB, etc. Adding obfuscation has however a production cost and can also trigger the "challenge spirit" of some attackers (some attackers will then try to hack the device for the challenge of it, just to show they can do it). It is thus recommended to only use it for devices already presenting a high security level.

Software

■ **Q8. Do you encrypt content stored on external memory?**

An attacker with physical access to a device might read external memory on the board and access sensitive content such as personal data and/or cryptographic material. An attacker might also modify the content of the external memory to fake data for instance.

■ **Q9. Is the cryptographic material unique for every device?**

If the devices have cryptographic material used to authenticate the device and secure the data (for instance keys), is it important for this material to be unique per device (see Q28).

■ **Q10. How is the cryptographic material generated?**

It is important the cryptographic material is generated in such a way it cannot be guessed by an attacker. In practice, it means using a TRNG or a PRNG with proper seed. An example of what **not to do** would be for instance to generate the device's key simply by hashing the device's ID.

■ **Q11. Are you using a secure cipher suite (strong encryption algorithms and strong keys)?**

An attacker could take advantage of a weak mode of operation to misuse the system (if for instance AES ECB is used, one could gain information, or try to perform replay attacks).

■ **Q12. Do you have Over-The-Air (OTA) firmware update?**

Providing an OTA firmware update mechanism allows to keep the devices up to date. It however needs to be implemented securely to prevent opening the door to attacks. The update must be transmitted securely, signed by an authorized trust entity and encrypted using proper encryption methods. The firmware must be checked by the device before proceeding to the update.

■ **Q13. Is the firmware update automatic?**

Without automatic update mechanism, devices' firmware will rapidly be outdated, allowing for known (and publicly available) vulnerabilities to be exploited by anyone, without requiring much knowledge.

■ **Q14. Does the firmware contain any sensitive data (e.g. Hardcoded credentials)?**

If it is the case, an attacker who gets his hand on a firmware file can retrieve the credentials and potentially compromise the ecosystem at scale.

■ **Q15. Are there any mechanisms to prevent against weak, null, or blank credentials?**

Such weak credentials allow an attacker to easily gain access to a device and/or service.

■ **Q16. Do you have mechanisms to isolate privileged code, processes, and data from portions of the firmware that do not need to access them?**

This prevents an attacker who can inject unauthorized code into a running "process" to access sections of the firmware that contain sensitive data (such as cryptographic material).

Communication

■ **Q17. Is the information transmitted over serial line protected?**

An attacker might want to access sensitive information and/or be able to fake the data sent by a device. Eavesdropping serial communications (such as communication between the microcontroller and a modem) requires very little knowledge and equipment, it is thus important to ensure that no sensitive information can be accessed/tampered with there.

■ **Q18. Is the technology/protocol used for wireless communications secure in your use case?**

Some protocols might have a secure configuration, but it might be impossible to reach such a configuration in your use case: for instance, while Bluetooth can be secure, if there is no way to establish a secure key exchange during pairing (using a pin for instance), there will be a risk of Man in the Middle attack (MitM).

■ **Q19. Are you implementing mutual authentication?**

Without mutual authentication, an attacker can potentially act as a MitM and steal information or inject fake data into the system.

■ **Q20. Are you exposing any credentials to internal or external traffic?**

Sending credentials over an insecure communication channel (for instance GPRS) could result in the credentials being intercepted by an attacker and further used to gain access to a backend server and/or service.

Infrastructure

■ **Q21. Is the access to the backend infrastructure protected?**

If possible, restricting access to the backend infrastructure is a good practice. For example, if devices always connect to a third-party communication provider's network to gain connectivity (Sigfox, or a mobile operator for instance), the backend infrastructure can be configured to only accept communications incoming from that network.

■ **Q22. If you have an OTA firmware update mechanism, is the update server secure?**

An insecure update server could result in device compromise at scale (an attacker uploading its own modified firmware for instance).

■ **Q23. Do you have any mechanism to detect a compromised device?**

A compromised device might behave outside the "expected range of operation". For instance, it could connect and send data at a different interval than the regular one, send data that should not be physically possible, etc. This can be detected in order to flag such devices or even disable their access.

■ **Q24. Do you have any mechanism to detect a rogue device?**

Attackers might have interest in introducing rogue devices to a system, i.e. unauthorized devices which pose as legitimate devices but might have enhanced and malicious capabilities.

■ **Q25. Is the backend infrastructure prepared to handle data flooding?**

An attacker who compromised one or more devices could start sending data to the backend at a much higher rate than expected. If the data is accepted by the backend, depending on the configurations, this could lead to data loss or extra costs (when using databases from a third-party cloud provider for instance). Having DDoS protection and load balancer is important to prevent this as well.

■ **Q26. Do you implement rate limiting?**

Without rate limiting, an attacker who gained access to a device could misuse it outside of its planned operation range.

■ **Q27. Do you have regular monitoring of the behavior of the devices in your ecosystem?**

Having such monitoring can help detect compromised devices that behave outside their defined operation range. It can be used to take actions such as disabling the device's accesses to the services.

Generic considerations

■ **Q28. What is the impact of a single compromised device on your system?**

Assuming one device is compromised by an attacker (for instance, the attacker got their hands on a device and extracted the firmware and cryptographic material), this can have severe consequences on the whole ecosystem if some cryptographic material is shared.

■ **Q29. Have you considered the operational consequences of one device being compromised? What about N devices?**

While a single device being compromised might not have a big operational impact, a few devices could start to have an operational impact depending on the use case.

■ **Q30. Are you collecting and storing only the minimum data required?**

In the case of a device collecting data, it is important to minimize the data collected and retained (especially in the case of personal data).

Additional resources

<https://www.enisa.europa.eu/publications/hardware-threat-landscape>

<https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot>

<https://github.com/OWASP/ASVS>

Glossary

JTAG

JTAG (named after the Joint Test Action Group which codified it) is an industry standard for verifying designs and testing printed circuit boards after manufacture.¹⁴ From a security perspective, it can be used by an attacker to retrieve a device's firmware and/or perform dynamic analysis on the device.

SWD

Serial Wire Debug (SWD) is an alternative 2-pin electrical interface to JTAG. It uses the existing GND connection.¹⁵

TRNG

True Random Number Generator (or a hardware random number generator (HRNG)) is a device that generates random numbers from a physical process, rather than by means of an algorithm.¹⁶

PRNG

A Pseudo Random Number Generator, also known as a deterministic random bit generator (DRBG), is an algorithm for generating a sequence of numbers whose properties approximate the properties of sequences of random numbers.¹⁷

MitM

A Man-in-the-Middle attack is a cyberattack where the attacker secretly relays and possibly alters the communications between two parties who believe that they are directly communicating with each other, as the attacker has inserted themselves between the two parties.¹⁸

DMA

Direct Memory Access is a feature of computer systems that allows certain hardware subsystems to access main system memory (random-access memory) independently of the central processing unit (CPU).¹⁹

RoT

RoT can be described as a set of implicitly trusted functions that the rest of the system or device can use to ensure security; it is the foundation on which a device maker can build their "tower of trust".²⁰

GPRS

¹⁴ <https://en.wikipedia.org/wiki/JTAG>

¹⁵ https://en.wikipedia.org/wiki/JTAG#Similar_interface_standards

¹⁶ https://en.wikipedia.org/wiki/Hardware_random_number_generator

¹⁷ https://en.wikipedia.org/wiki/Pseudorandom_number_generator

¹⁸ https://en.wikipedia.org/wiki/Man-in-the-middle_attack

¹⁹ https://en.wikipedia.org/wiki/Direct_memory_access

²⁰ <https://www.psacertified.org/blog/what-is-a-root-of-trust/>

General Packet Radio Service (GPRS) is a packet oriented mobile data standard on the 2G and 3G cellular communication network's global system for mobile communications (GSM).²¹

DDoS

A distributed denial-of-service attack (DDoS attack) is an attack in which the perpetrator(s) seek to make a machine or network resource unavailable by flooding it of requests. The incoming traffic flooding the victim originates from many different sources.²²

Server & Cloud Security Checklists & Requirements

Checklist for security in water process control networks

This checklist is a translated and slightly updated version of the one provided by *Jaatun, Røstum and Petersen* [35]. The checklist is also available as an Excel spreadsheet with automatic tallying, as described by *Jaatun et al.* [35], and illustrated in Figure B.1.

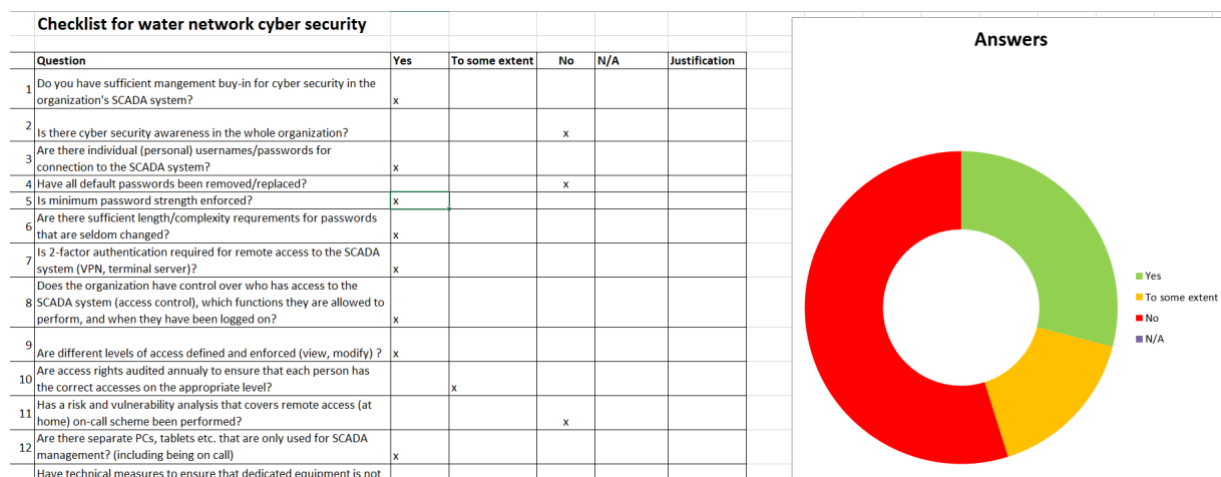


Figure B.1 Filled-in checklist with high-level overview

Q1. Do you have sufficient management buy-in for cyber security in the organization's SCADA system?

Q2. Is there cyber security awareness in the whole organization?

Q3. Are there individual (personal) usernames/passwords for connection to the SCADA system?

Q4. Have all default passwords been removed/replaced?

²¹ https://en.wikipedia.org/wiki/General_Packet_Radio_Service

²² https://en.wikipedia.org/wiki/Denial-of-service_attack

- Q5. Is minimum password strength enforced?
- Q6. Are there sufficient length/complexity requirements for passwords that are seldom changed?
- Q7. Is 2-factor authentication required for remote access to the SCADA system (VPN, terminal server)?
- Q8. Does the organization have control over who has access to the SCADA system (access control), which functions they are allowed to perform, and when they have been logged on?
- Q9. Are different levels of access defined and enforced (view, modify)?
- Q10. Are access rights audited annually to ensure that each person has the correct accesses on the appropriate level?
- Q11. Has a risk and vulnerability analysis that covers remote access (at home) on-call scheme been performed?
- Q12. Are there separate PCs, tablets etc. that are only used for SCADA management? (Including being on call)
- Q13. Have technical measures to ensure that dedicated equipment is not used for other things been enforced? (Restriction on software that can be installed, access to internet blocked, etc.)
- Q14. Is there separation and/or security barriers between administrative and SCADA networks?
- Q15. Is there separation and/or security barriers between internet and SCADA networks?
- Q16. Has the waterworks considered criticality/importance of the various outstations/water installations?
- Q17. Are there separate SCADA networks for water and wastewater, respectively (separate zones)?
- Q18. Are SCADA networks (logically) divided in geographical zones?
- Q19. Are there mechanisms that restrict the possibility to access other parts of the SCADA network from a given out-station?
- Q20. Is there a system for registering incidents and deviations in the organization?
- Q21. Is there a separate system for systematic registration of SCADA incidents and deviations (failures)?
- Q22. Are there documented routines for handling of SCADA incidents and deviations?
- Q23. Is it easy for non-authorized personnel to get access to information on geographic placement of water infrastructure?
- Q24. Is there redundancy of critical systems/components (two servers, communication, power supply, server room)?
- Q25. Can all installations be operated in manual mode if connection with remote installations (water treatment, ...) is lost?
- Q26. Are there contractual agreements for cyber security with external service providers (contractors, suppliers, ...)?
- Q27. Have intrusion detection systems (IDS) been deployed?
- Q28. Are there antivirus routines/mechanisms in the SCADA system?
- Q29. Are there routines for updating/patching of SCADA system components?

Q30. Has the water network operator performed an analysis of SCADA network connection to other networks (Internet, administrative systems etc.) that could represent a threat?

Q31. Do contingency plans also cover SCADA incidents?

Q32. Have contingency plans been updated in the last 12 months?

Q33. Is there sufficient external physical protection of water infrastructure so that it is not too easy to get physical access to the infrastructure?

Q34. Is the physical protection of the installations commensurate with the importance of the water installations?

Q35. Are there good routines for ensuring sufficient training and continuing education of personnel regarding Cyber security?

Q36. Are periodic risk analyses that also cover cyber security and SCADA performed by the organization?

Q37. Do the emergency preparedness exercises of the organization also cover cyber/SCADA incidents?

Q38. Does the organization have preparedness and prepared actions for alternative operation of the water/wastewater system in case of failure in the whole or part of the SCADA system?

Q39. Does the organization have sufficient internal competence in cyber security and SCADA, or is it wholly or in part dependent on vendors and consultants?

Q40. Does the organization have an agreement concerning a full backup of the SCADA configuration?

Q41. Does the organization have subject-matter experts with sufficient procurement competence when ordering a new SCADA system?

Q42. Are there up-time requirements for the SCADA system?

Q43. Are all system solutions and SCADA specifications well documented?

Q44. Is periodic updating of system/SCADA documentation performed?

Q45. Are periodic tests of the SCADA system to check that the installation works according to intentions performed? Ref. internal control that should include checking that, e.g., measurement values reported are correct, and that algorithms for controlling operations are correct and according to plan.

Cloud Security Requirements for Critical Infrastructure

These cloud security requirements are derived from a number of good-practice publications as described by *Bernsmed, Meland and Jaatun* [37], and updated by *Røstum and Jaatun* [28].

Requirements related to data storage

Encryption: Ensure that data is not stored in clear text when not in use

- All data is encrypted when stored. Disk encryption is sufficient (including virtual disks).
- Data from each infrastructure operator must be encrypted with separate encryption keys.

Physical location: Ensure data is stored in a specific geographic location

- As a rule, cloud providers based in Europe should be preferred.
- Consider national archival legislation – additional copies for some data may have to be stored within national borders.

Isolation: Ensure that the customer retains ownership of own data

- All data stored in the cloud solution remains the property of the critical infrastructure provider.
- A data processing agreement shall be entered into with the supplier. This can be with a third party that develops services using a cloud service provider, or directly with the cloud service provider itself.
- The cloud service provider may not use data from the critical infrastructure provider for the former's own purposes.
- The customer retains ownership of its data also after the termination of the service contract.

Portability: Ensuring portability of customer data

- Data must not be locked in on the cloud provider's platform but must be exportable to a pre-agreed (preferably open) format on demand.

Integrity: Ensure correctness and consistency in customer data

- Integrity must be maintained for all data stored in the cloud solution.

Deletion: Ensure proper deletion of all data upon customer request

- All replicated data shall be deleted within a specified deadline when requested by customer. (Refer to GDPR for privacy-related requirements).

Backup: Ensure backup is performed and maintained in a proper manner

- The cloud solution shall create backups at least [daily]
- A local (off-cloud) backup of the cloud data shall be performed at least [weekly] - this should be usable also when the cloud service is not available.
- If a risk analysis has been performed, and the availability of the provider has been assessed as sufficient, the remaining risk may be accepted, and the requirement of local mirroring may be dropped.
- A detailed scheme for how long backups should be retained must be devised. E.g.:
 - daily backup: 21 days;
 - weekly backup: 12 weeks; (also applies to local mirror if present)
 - monthly backup: 6 months; (last Friday of each month)
 - quarterly backup: two years (last Friday in January, April, July and October)
- Backups stored in the cloud must be checked by restoring to shadow system at least monthly. Risk analysis may conclude that frequency must be increased. Agreements must be made with service provider regarding how this can be documented.
- Local (off-cloud) backup must be checked [weekly] (when it is created)
- The cloud provider shall commit to a guaranteed maximum time for restoring of backup copy.
 - If there is a local mirror, it should be available when it is created. Cloud provider should cater for restoration from local copy if the local copy is the most recent.
 - If there is no local backup, the guarantee relates to the restoration of the cloud backup.
- Backup copies must be stored geo-redundant with respect to where data are normally stored

Requirements related to data processing

Isolation: Ensure that all data is isolated from other customers' data

- All in-memory data shall be segregated from data belonging to other customers
- The cloud provider must implement mechanisms that ensure that different virtual machines do not influence each other
- Data sent to the cloud service related to a specific request are not visible to other users of the service

Monitoring: Ensuring that breaches of permissible use agreements are detected

- Behaviour of running virtual machines (VMs) shall be monitored continuously

Physical location: Ensure data is processed in a specific geographic location

- As a rule, all data should be processed in data centres based in European Economic Area.

Migration: Ensure that migration between different physical servers is performed securely

- All VMs must be encrypted during migration

Requirements related to data transfer

Encryption: Ensure that data is not transferred in clear text

- Up- and downloading of data to/from the cloud service must be encrypted.
- All communication stages should be encrypted.
- End-to-end encryption shall be used whenever possible.

Integrity: Ensure correctness and consistency in customer data

- Up- and downloading of data to/from the cloud service must be integrity protected.

Isolation: Ensure that all data is isolated from other customers' data

- The cloud service provider must offer network isolation between customers, ensuring that no data traffic to/from one customer can be eavesdropped on by another

Requirements related to access control

Access control for administration: Ensure secure access to the cloud administrative interface (dashboard)

- The cloud provider shall enforce a good practice password policy, focused on length and complexity of passwords
- The cloud provider shall support multi-factor authentication
- The cloud provider shall support third-party authentication solutions for simple login (SAML/OpenID)

Access control for users: Ensure secure access for cloud users

- The cloud provider shall provide a system for creating, updating, suspending and deleting user accounts, to remove access of employees when they leave the organization
- All cloud users should have unique user accounts; no joint accounts are to be used
- Access to cloud services should be role-based
- On need it shall be possible to separate data that should not be displayed on certain platforms (portable devices, etc.)

Physical access control: Ensure that data centers are secured in an acceptable manner

There may be big differences between cloud providers. Big providers typically have good physical access control, so these requirements may be more relevant for smaller providers.

- Datacenters must be protected by physical security perimeters (fences/guards/surveillance/locks)
- Secured areas must be subject to surveillance, and access must be restricted to authorized personnel
- The provider's personnel who have access to customer data have been subject to background checks (e.g., security clearance).
- Access to applications or source code must be restricted to authorized personnel with “need to know”.

Requirements related to security procedures

Audit: Ensure that cloud services can be audited

There may be big differences between cloud providers, Big providers can be expected to have regular external audits. Small providers may need to carry out audits on demand.

- The cloud provider must perform regular audits of the cloud service.
- The customer must be given access to an independent audit report that documents the security of the cloud service.
- The audit should cover physical, technical, and organizational security measures, and document whether these are suitable for the cloud service in question.
- In case of composed cloud services, the assessment should include assurances that the security of each component in the provider chain involved in the processing of the customer's data satisfies the security requirements.

Countermeasures: Ensure that the cloud service implements defense mechanisms

This is important to ensure uptime of the system.

- Firewalls are installed and configured.
- Mechanisms to protect against or minimize the impact of DoS/DDoS are installed and configured.
- Mechanisms for data loss prevention are implemented.
- All software is updated with the latest security patches. Responsibilities must be clarified – will differ based on what service model is chosen.
- The service is established with a redundant configuration.

Testing: Ensure that the security of a cloud service can be tested.

- The provider should perform regular vulnerability scans.

Detection: Ensure that attacks and intrusions will be detected.

- An Intrusion Detection System (IDS) has been installed and configured correctly.
- Disks, memory and network is regularly scanned for malware.
- Procedures for monitoring and regular audit of logs have been established.

Alerts: Ensure that customers get information about security related changes

- Provider must inform their customers of any vulnerabilities that are discovered. Should be included in contracts/SLAs. Deadlines for alerts should be specified in SLAs.
- Provider must be able to provide information on security mechanisms and control mechanisms that are in place.
- When provider has issued an alert on a vulnerability, it must subsequently inform customers of which changes have been performed (patches applied, etc.).
- Customers must be notified of changes in geolocation (where the cloud server /datacenter physically resides).

Restore: Ensure that the cloud service can be restored after an attack

- The provider must maintain periodic checkpoints with the state of virtual machines.
- The provider must guarantee that if the system is compromised, the service can be re-established in a secure fashion within a specified time interval. This should be specified in the SLA.
- The provider must regularly (at least once a month) perform tests on restore from backup.

Key management: Ensure secure management of cryptographic keys

- The provider must be able to document a secure key management scheme on demand.

Transparency: Ensure transparency of the cloud service and its security mechanisms

- The provider must on demand be able to demonstrate/explain (but not necessarily provide/hand over documentation on) architecture, solution design and security mechanisms.

<i>Requirements related to incident management</i>
--

Response: Ensure that there is a system in place for responding to security incidents

- The provider should provide periodic summaries that show the level/amount of incidents (e.g., number of DoS attacks previous period)
- Incidents should be categorized according to a well-defined pattern.
- The SLA should contain requirements with deadlines for when the provider will respond to a given category of incidents.
- The provider must have a documented process for handling security incidents, preferably based on an international standard or guideline.

Logging: Ensure that the provider logs security events

- The provider must store audit logs with registered activity for privileged uses, authorized and un-authorized access attempts, anomalies, and information security events.

- The provider must on demand be able to produce an event log relevant for the individual customer.

Reporting: Ensure that the provider reports security incidents to their customers

- The SLA must contain deadlines for when security incidents must be reported to customers.
- Alerts should be given accordance with a predefined alert procedure and communication channel.
- In case of incidents, the provider must estimate time to repair/recovery. If estimates prove to be overly optimistic, new information must be given.

Investigations/forensics: Ensure that the provider can offer necessary support to an (official) investigation

- The provider should have guidelines for how evidence can be gathered and handed over to relevant authorities in case of security incidents connected with crimes or other legal issues.

Requirements related to aggregated services

Outsourcing: Requirements for third-party providers

- The provider must only outsource services (or parts of services) to third party providers that process and store data in EU/EEC

Surveillability: Ensure the customer's ability to inspect how the service is composed

- The customer should have the capability to inspect which components the service is composed of.
- In case of changes in composite services, the provider must inform the customer about the changes (e.g., new subcontractor)

ML & AI Security Checklist

This checklist is based on the 10 security principles applied to Machine Learning as presented in *An architectural risk analysis of machine learning systems: Toward More Secure Machine Learning* [28].

Q1. Have you identified and secured the weakest link in your system? (Secure the weakest link)

If you put yourself in the shoes of an attacker, where would you attack your system? Did you ensure the algorithm you are using aren't malicious?

Q2. Do you have several levels of security in place? (Practice defense in depth)

For instance, do you secure the training data (authentication and authorization)? Do you in addition anonymize the dataset to prevent information leaks? Are you limiting users' access?

Q3. Does your system exhibit any insecure behavior on failure? (Fail securely)

Are you exposing the confidence score (which could be exploited to craft adversarial inputs)? Are you transmitting low confidence results to untrusted users? As a rule of thumb, consider carefully what information are reported to the users.

Q4. Do you grant only the minimum required access to users? (Principle of least privilege)

Has the system still access to the (potentially sensitive) training data (in the case of an offline model)? Are you removing the access when no longer required?

Q5. Do you keep components separate as much as possible? (Compartmentalize)

Having distinct components with clearly defined role helps applying security principles. Also, if one of the components gets compromised, the others should not be impacted.

Q6. Are you using the simplest approach to solve the problem? (Keep it Simple)

Complexity in a system leads to greater risk of bad and buggy implementation, leading to potential security issues. Using an overly complex algorithm which is difficult to understand might also hide those potential security issues to someone who doesn't fully understand it.

Q7. How do you handle privacy? (Promote Privacy)

Are you using techniques such as differential privacy to ensure an attacker cannot leak any information?

Q8. How do you protect confidential information? (Hiding secrets is hard)

Do you anonymize data? If so, do you ensure it cannot be deanonymized?

Q9. Are you trusting any external sources? (Be reluctant to Trust)

Where are you getting your data from? Can you trust that source? Where are you performing your computation? Can you trust that machine?

Q10. Are you relying on community resources?

Do you use any existing model (to perform transfer learning for instance) which could be used to attack your model? If you are using public datasets, can they be trusted as authentic and of quality?



Leading urban water management to its digital future

digital-water.city
 **digitalwater_eu**



digital-water.city has received funding from the European Union's H2020 Research and Innovation Programme under Grant Agreement No. 820954.