IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 December 2022, accepted 23 January 2023, date of publication 1 February 2023, date of current version 7 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3241552

RESEARCH ARTICLE

Thinking in Systems, Sifting Through Simulations: A Way Ahead for Cyber Resilience Assessment

FRANCESCO SIMONE, (Member, IEEE), ANTONIO JAVIER NAKHAL AKEL, GIULIO DI GRAVIO, AND RICCARDO PATRIARCA

Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, 00184 Rome, Italy

Corresponding author: Francesco Simone (francesco.simone@uniroma1.it)

ABSTRACT The interaction between the physical world and information technologies creates advantages and novel emerging threats. Cyber-physical systems (CPSs) result vulnerable to cyber-related disruptive scenarios, and, for some critical systems, cyber failures may have fallouts on society and environment. Traditional risk analysis in no more sufficient to deal with these problems. New techniques are gaining increasing consensus, especially those based on systems theory. In this context, the System-Theoretic Process Analysis for Security (STPA-Sec) extends the Systems-Theoretic Accident Modelling and Processes (STAMP) model considering cyber threats, and identifying unsafe and unsecure controls throughout a cyber socio-technical system. Despite its large usage as a descriptive tool, there is still limited use of STPA-Sec in (semi-)quantitative terms. This article presents System-Theoretic Process Analysis for Security with Simulations (STPA-Sec/S), a methodological interface between STPA-Sec and quantitative resilience assessment based on simulation models. The methodology is instantiated in a demonstrative case study of a water treatment plant, and its critical CPSs which may impact both community health, and environment. The obtained results show how STPA-Sec/S foster systems understanding, allow a systematic identification of its major criticalities, and the respective quantification.

INDEX TERMS Cyber security, cyber-socio-technical systems, hazard analysis, industrial systems engineering, resilience management, systems modeling.

I. INTRODUCTION

Over recent years, the call for digitalization and automation has demanded an increasing attention towards human-machine interactions [1]. The cooperation between human agents and smart and interconnected devices stresses the need to acknowledge a joint social and technical dimension [2]. Starting from the 1950s notion of socio-technical systems [3], cyber-socio-technical systems are here used to emphasize the cyber-physical integration with human related elements [4]. On one hand, modern industrial systems that are more prone to human slips and lapses might benefit from this transformation, on the other, the same systems might suffer from unexpected new threats and disruptions. These latter emerging as a result of the tight interactions between the physical world and the Information Technology (IT) environment. A cyber security issue does not only refer to data or information leakage anymore, but it can have tangible consequences, too. In this context, an update of risk management practices becomes fundamental for safety and security purposes. Accident models support the definition of causal factors leading to an accident, and hence, they support the identification of necessary measures to be implemented in order to avoid future or similar consequences and/or reduce their likelihood [5], [6]. Due to the increment of systems complexity, many accidents do not simply result from one – or a set of – trigger event(s), but they are caused by more complex intertwined etiological structures [7]. These accidents involve many different factors that increase variability of normal systems' operations, such as human factor, mission, smart devices, financial aspects, and information exchange [8]. Complex systems are characterized by nonlinear behaviors generated through the interactions among the system components, stressing the need to develop systematic methods to manage safety and security

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Merlino.

risks simultaneously [9]. Dysfunctional interactions among the system components might be a suitable way to describe accidents [10], [11], and such complexity-oriented accident analysis models seem necessary. They may rely on systems theory which stresses the focus on both the system operations and the management process related with the analyzed system itself [12].

The systems thinking techniques consist of three main aspects: (i) the attributes of system elements; (ii) the interconnections among elements; (iii) the functional purpose of the system. Systems theory can be applied in risk, safety and security management to analyze the interactions among system components and the overall system behaviors [13]. On this path, an interesting stream of research has been built upon the Systems-Theoretic Accident Modelling and Processes (STAMP) model, which is rooted in both control theory, and previous experiences with hierarchical safety control actions [12], [14]. Based on STAMP, a powerful accident analysis tool called System-Theoretic Process Analysis for Security (STPA-Sec) has been proposed in [15]. STPA-Sec provides a methodology to perform hazard analysis suitable for both physical and cyber accidents, especially in complex socio-technical cases.

In this context, this paper presents a novel methodology for cyber socio-technical systems modelling specifically suited for cyber threats analysis. In this regard, this work enhances qualitative system-theoretic approaches by adding a procedure to allow quantifying their output. The methodology integrates systems theoretic modelling with simulation tools to enhance process engineering design and practice via cyber resilience assessments. To this purpose, cyber resilience is defined as the ability of the system to anticipate, withstand, recover from, and evolve to improve capabilities in the presence of cyber threats [16]. In operational terms, the proposed methodology, STPA-Sec/S (System-Theoretic Process Analysis for Security through Simulations) relies on STAMP modelling, and extends a STAMP-like technique, to calculate cyber resilience metrics in a simulative environment.

The proposed methodology is instantiated in a case study of an hypothetical water treatment plant (specifically a Sea Water Reverse Osmosis plant) as a significantly critical segment of a water supply system. Digitalization strongly improves operations and systems' performance ensuring higher efficiency and coordination, and such benefits acquire particular relevance in critical infrastructure systems and in their related industrial settings. In this sense, the water supply systems represent a prime example. Sensing instrumentation, communication networks, and computing and control algorithms are, by now, jointly integrated within the water supply systems to enhance their operations. A successful cyber attack against a water system may result in water shortages, but also in contaminated – may be harmful – water supply, or even in potential environmental contamination. For this reason, the water sector strongly demands dealing with the potential vulnerabilities due to cyber-related failures and their consequent disruptive scenarios. Recent events demonstrate the potable water sector to be extremely vulnerable under this point of view, demanding for solutions in this sense. For example, two cyber-attacks were conducted against water distribution system in Israel during 2020 creating an ''unpredictable risk scenario'' [17]. Even though no consequences have been declared by the responsible authorities, there was an open chance that thousands of people would have been fed with low quality – may be poisoned – water or left without it. These attacks were not isolated, as observable by multiple similar events distributed all over the world [18]. On this basis, we believe the case study represents a priority domain to be investigated from a joint safety, security and environmental perspective.

The remainder of the manuscript has been organized as follow. Section II reviews literature on the uses of system-theoretic approaches in different domains. In Section III, STPA-Sec/S, our novel methodology, is presented, discussing the integration between system-theoretic approaches with cyber resilience assessment based on simulations. The methodology is instantiated in Section IV for the case study at hand, presenting results and discussing their validity for operative cyber risk management. In Section V concluding remarks with possibilities for future domain of applications, and ideas for further development, are provided.

II. LITERATURE REVIEW

Traditional system safety approaches are being challenged by the introduction of new technologies and the related increasing complexity. The dependencies, the relationships, and the interactions between systems parts make it difficult to assess and control system functioning in a linear mechanistic way. These concepts have been previously introduced into Prof. Leveson's STAMP, i.e., System Theoretic Accident Model and Processes, CAST, i.e., Causal Analysis based on STAMP, and STPA, i.e., Systems-Theoretic Process Analysis [12]. While the STAMP is a descriptive model of the system under investigation, CAST and STPA permit to analyze past (CAST) and probable future (STPA) accidents and incidents.

STAMP and its nested techniques are widely applied in domains dealing with socio-technical systems, as the aviation sector. For example, in [19] authors propose a defect prediction model for radars' software based on STAMP theory; in [20] a mid-air accident is analyzed through CAST, also expressing the interactions between people, technical equipment, and environment; in [21] STAMP and STPA are used at first to identify unsafe scenarios encompassing both technical and organizational aspects of a flight demonstrator, and then to guide the proposal of safety control measures. The space sector has been shown to have benefitted from STAMP as well, e.g., in [22] STAMP is used to map human-machine interactions during various stages of the lifecycle of the Apollo system. In [23] the authors developed a CAST analysis to describe the International Space Station EVA 23 water intrusion incident in order to explore complex interconnections, and real-time flight organizational operations, pairing

it to safety recommendations. Other exemplary domains of application comprehend: the healthcare services [24], the automotive industry [25], the transportation systems [26], the nuclear energy generation [27], and the process industry [28], among others.

The traditional risk perspectives are not necessarily adequate to deal with cyber threats. As per the context of this paper, special attention should be devoted to the application of system-theoretic approaches oriented towards cyber-related failures. In [29] STAMP-Sec is introduced as an extension of the STAMP method to model security incidents as the result of inadequate controls rather than strictly failure events. Cyber threats were then translated into security constraints that can inform the design of security-critical systems. An extension of the STPA method has been proposed in [30], namely, the System Theoretic Process Analysis for Security (STPA-Sec). STPA-Sec shares the same principles with its traditional safety-oriented counterpart, STPA, although the results and detailed procedures are slightly different. The security-tailored methodology is also tested on a nuclear plant reactor showing the set of conditions which could lead the plant to a loss. STPA-Sec has been successively applied in a variety of sectors. In [31] STPA-Sec is used to guide the choice of security requirements for the design of a drone. The analysis is made upon three different levels of detail to provide traceability to the system owner's mission, to make systems security easily understandable. In [32] STPA-Sec is applied to understand and to elicit systems security requirements during the conceptual stage of development of a space system. The obtained results provide insights in getting viable systems security requirements in terms of traceable security, safety, and resilience. STPA-Sec has been also used (e.g.) to understand security and resilience requirements in the early stages of the development of a refueling aircraft [9], in the design of the controls of a smart electrical micro-grid [33], or to perform a preliminary hazard analysis, to evaluate available alternatives, and to ensure safe operations of an autonomous driven vehicle [34].

Although the usefulness of these techniques has been widely proven, a recent literature review on STAMP-like techniques documented the increasing tendency by scholars to complement them with other approaches [13]. For example, in [35] authors proposed a new accident causation theory based on STAMP and on a risk management framework to go beyond human, organizational and technological characteristics encompassing sociological factors (legislative, regulatory, and cultural). While STAMP and its related tools have a purely qualitative nature, other research is aimed at their integration with risk and losses quantification methods, e.g., [36]. A shared solution relies on the usage of model checking techniques to improve or verify system requirements against their actual configuration [37], [38], to guide prioritization of hazardous scenarios highlighted by the STPA analysis [39], or to perform a safety assessment providing a formal and unambiguous representation of the system and its related threats [40]. In [41] instead, authors proposed an integration of the STPA and the Functional Resonance Analysis Method (FRAM) to carry out a safety analysis identifying potential risks and providing mitigation measures. The FRAM model is verified against the STPA-based safety properties and used as a starting point for a quantitative model checking.

Also, the system dynamics modelling is used in [42] to map the human factor, and to identify the impacts of avionics reconfigurations during system development, operations, and revision. A hazard analysis made through STPA become the foundation of this study. An extension of STPA is also provided in [43] where it is proposed a new approach named STPA-RAM (Reliability, Availability, and Maintainability). This latter consists in the utilization of a discrete event simulation to transform the feedback control loops into a set of stochastic Petri nets.

A quantitative resilience assessment may complete the analysis, too. Accordingly, in [44] authors proposed a methodology to carry a quantitative resilience assessment based on STAMP modelling of the system under analysis. Assessing the system resilience under the influence of a disruption, demand for two main steps: (i) development of the STAMP model of the system permits to describe system relationships, (ii) modeling parameters and a quantitative resilience metric. Also, STPA is used to identify system hazards and accidents, determine unsafe control actions, and find out their causes. The proposed method is also tested through an application on a diesel oil hydrogenation system.

Besides multiple developments to complement STAMP/STPA methodologies, there is still little – but recent – track of proposal integrating STPA-Sec. In [45] this latter is integrated with the Combined Harm Analysis of Safety and Security for Information Systems (CHASSIS) methods for the information lifecycle analysis to complement and generate additional considerations on top of the ones provided by STPA-Sec analysis. An additional methodological structure is provided also in [46] where the NIST (National Institute of Standards and Technology) requirements have been integrated throughout the security analysis. In [47] it is pointed out how STPA-Sec lacks in considering the IT security issues such as data confidentiality. To overcome this gap, STPA-DFSec (DF stands for data flow) is presented. This new framework introduces a data-flow diagrams for information security considerations. A study on a vehicle digital key system is shown to instantiate the modified approach and compare it with the classic STPA-Sec methodology. In [48] STPA-Sec becomes STPA-Priv which relaxes the assumptions to consider only closed loop controls and extend the analysis also to open-loop controls. In [49] the authors proposed an ontology-based technique that extends STPA-Sec placing emphasis on the security threat scenarios and their identification.

Nonetheless, if the presence of some quantitative analyses improving classic STPA has been shown, there is still no track of guidelines for the development of quantitative approaches making the STPA-Sec assessment less expert-dependent. The

current research in system-theoretic based cybersecurity still focus in improving the cyber issues identification. In this paper, we propose a novel approach, i.e. STPA-Sec/S, which integrate the STPA-Sec analysis with a quantitative resilience assessment based on simulations. The STAMP model is used to guide the simulation model development, then STPA-Sec helps identifying system cyber vulnerabilities which constitutes the causal scenarios to be simulated. We rather aim to provide a methodological guide to assist the model development, the system cyber resilience assessment and the system's criticalities evaluation.

III. METHODOLOGY

This section describes the theoretical aspects and the operational implications of our novel methodology, namely, STPA-Sec/S. At first, the theoretical fundamentals of both STPA-Sec and simulation techniques in the context of socio-technical systems are introduced. The description is then complemented to show how to perform the novel integration, including guidelines to translate the elements of the STAMP model into the simulation environment. It is shown how to frame simulation scenarios based on STPA-Sec outcomes. The resulting resilience metric definition relies on system-theoretic approaches, too.

A. SYSTEM-THEORETIC PROCESS ANALYSIS FOR SECURITY (STPA-SEC)

STPA-Sec is a hazard analysis technique based on an extended model for threats against IT systems under attack. It consists of an early concept analysis to identify inadequate safety controls in system design. Its aim is to assist security and safety management in the definition of the requirements and the countermeasures against cyber attacks [15]. The method can be used to analyze complex system interactions between their components, or throughout organizational levels. STPA-Sec comprehends four phases:

- *Purpose of the analysis.* STPA-Sec requires a definition of systems boundaries and its hazards. Five sub-steps are needed to precisely define the constraints and the scope of the analysis. (i) Problem framing, to identify which elements must be protected and to define how these elements could be protected. During the problem framing phase, a problem statement is synthesized. (ii) Losses identification, in terms of the values which impact the stakeholders (financial losses, productivity losses, social impacts or reputational damage, legal liability). The losses must be defined by identifying the stakes (goals, aims, values, or missions) defined by the stakeholders. The stakes are transformed into system losses to be avoided in the system security analysis. (iii) System hazards identification, which provides the definition of components to be analyzed and their boundaries. This allows to highlight the difference between hazards and losses. Thus, the hazards can be defined by a combination of conditions that, in a particular situation, will lead to a security loss in the system. (iv) System safety and security constraints identification, which are the security conditions to satisfy, to prevent, or reduce the system hazards. (v) Refine systems hazards. When a hazard comprehends more than one sub-hazard. This step may be suitable for large analysis and complex applications. In such cases, creating a new sub-hazard with its specific safety constrains might be appropriate.
- *Safety Control Structure (SCS) model.* The system is modelled through a hierarchical structure, leading to the development of a STAMP model. This enables mapping the interactions among the system components. The model is composed by control loops (i.e., control actions and feedbacks), and inputs/outputs parameters, on which the controllers act.
- *Unsafe and Unsecure Control Actions (UCAs) identification.* UCAs are control loops which lead to hazards or losses when ad adverse condition verifies. STPA-Sec proposes four possible scenarios describing unsafe or unsecure control loops: (i) a not provided feedback or control action which leads to a hazard; (ii) a provided feedback or control action which leads to a hazard; (iii) a feedback or a control action provided too early, too late, in wrong order or with an inappropriate application; (iv) a feedback or a control action provided (or stopped) for too long, or too short.
- *Security-related causal loss scenarios identification.* Following the definition of the UCAs, it is possible to describe the causal factors that may lead to the unsafe or unsecure actions, and in turn leading to unacceptable system losses. STPA-Sec foresees several types of loss scenarios to be considered in the analysis. A first type of scenario describes why an UCA would occur. A second type, defines why the feedback or the control actions would be improperly or not executed. Lastly, the scenarios in which UCAs lead to unacceptable losses. To retrieve this type of scenarios, one can move backward from an UCA looking for what could induce the controller to generate (or to receive) that unsafe control action (or feedback). Therefore, the scenarios can be related to: (i) controllers, (ii) system behaviors, (iii) control actions, (iv) context.

B. MODELLING AND SIMULATION FOR CYBER RESILIENCE ASSESSMENT

To quantitatively evaluate cyber resilience, it is necessary to experience faults and damages caused by cyber attacks. In this regard, it is often difficult to carry on experiments, since their impact on critical systems may have disastrous consequences. This issue is by-passed using simulation techniques, which make it possible to assess system behavior without producing real disservices. The use of model-based approaches for quantitative cyber resilience assessment is a common practice (e.g., [50], [51], [52], [53]). These approaches consist of modelling all the phases of the process of interest using software, and then simulate system's failures to finally compute

certain resilience metrics. Common modelling and simulation methodology to assess cyber resilience can be resumed in the following steps [54]:

- *Problem identification.* To identify a problem statement containing the information on the system to be analyzed, the knowledge of its acceptable states, and a hypothesis of one (or a set of) successful cyber attack event(s).
- *System description.* To define the system's boundaries and the simulation testbed. The selected system is described breaking down it to its elements. For cyber resilience evaluation, elements description must consider both physical and IT components, along with control strategies and communication aspects.
- *Digital model design.* Elements and their functionality are formalized through adequate analytical relationships, which allow transferring the digital model into a selected simulation environment. The model is then validated to obtain a representative nominal functioning scenario.
- *Metrics definition.* To define indicators to study the cyber resilience problem. Resilience indicators rely on chosen definition of cyber resilience and the critical variables of both the physical systems and its digital counterpart.
- *Modelling failure scenarios.* Cyber attacks must be modelled on the simulation environment taking into account attack's characteristics. System response and recovery capacities are modelled accordingly.
- *Cyber resilience assessment.* Each cyber attack generates a disrupted performance curve which has to be compared against system nominal performance. Cyber resilience is then calculated accordingly.

C. INTEGRATING STPA-SEC AND SIMULATION: SYSTEM THEORETIC PROCESS ANALYSIS FOR SECURITY WITH SIMULATIONS (STPA-SEC/S)

Relying on the theoretical aspects proposed by STPA-Sec and simulation techniques, Fig. 1 depicts the steps to develop a cyber security analysis through the System Theoretic Process Analysis for Security with Simulation (STPA-Sec/S), which incorporates analytical steps from both methodological instruments. Accordingly, the steps of STPA-Sec/S are:

- *Step 1: Define the purpose of the analysis.* The first step consists in defining the system at hand, its mission and operating scenarios, and to frame the problem to be investigated. Subsequently, the aim of the system mission, its related losses, the hazards, and the safety constrains shall be identified for each operating scenario.
- *Step 2: Safety control structure modelling.* A safety hierarchical structure is required to map the system process. Therefore, the Systems-Theoretic Accident Model and Process (STAMP) model is used to create a model for the system at hand mapping the interactions among its elements. The STAMP model reports the control actions and the feedbacks to monitor and manage the controlled process.
- *Step 3: Identify unsafe/unsecure control actions*. At this stage, based on the STAMP model, it is possible to identify those loops that – under certain operating conditions – may lead to hazards or losses. Besides, the control actions or the feedbacks could be unsafe or unsecure when: (i) not providing the feedback or the control action leads to an hazard; (ii) providing the feedback or the control action leads to an hazard; (iii) providing the feedback or the control action too early, too late, in wrong order, or with an inappropriate application leads to an hazard; (iv) the feedback or the control action provided are stopped too late, or too soon, leading to an hazard. Step 4 is proposed to substitute the causal factors identification.
- *Step 4: Digital model design.* The identification of the critical elements guides the digital model design. The model boundaries, in terms of the features and the relationships to be reproduced, strictly depends on which part of the safety control structure model appears to be critical from Step 3. Each critical element, its related entities, and its linked relationships shall be reproduced in the simulation environment. For this purpose, the STPA-Sec/S includes guidelines to ensure a coherent digital model conversion from STAMP (see Section III-C1). Model validation is included in this step to ensure a representative nominal scenario [55].
- *Step 5: Define resilience metrics.* At this stage, one (or a set of) system performance cyber resilience metric(s) shall be defined. The definition of the metrics relies on the specific purpose of the analysis (i.e., which system losses are considered, which are the hazard causing the losses), and on the capabilities of the developed digital model. These metrics lay the foundations for the subsequent performance analysis and as such they shall be SMART, i.e. specific, measurable, attainable, realistic and tangible [56].
- *Step 6: Model faults and effects.* To proceed with the cyber resilience assessment, it is necessary to reproduce the system failures caused by cyber attacks with their consequent effects, as well as system restoration capacity, if any. The failure scenario definition is based on the outcomes of the STPA-Sec analysis: scenarios to be simulated are those arising from the occurrence of UCAs. The attacks' characteristics must be modelled to generate representative simulation outputs.
- *Step 7: Perform resilience assessment.* The cyber resilience quantification analysis is finally performed by looking at the simulation outputs in terms of the pre-defined performance metrics. Simulations could reproduce a single specific failure scenario by fixing attack characteristics, or they could be run multiple times (e.g., by considering stochastic parameters and perform Monte Carlo simulations) to define system behaviors under different attack parameters.

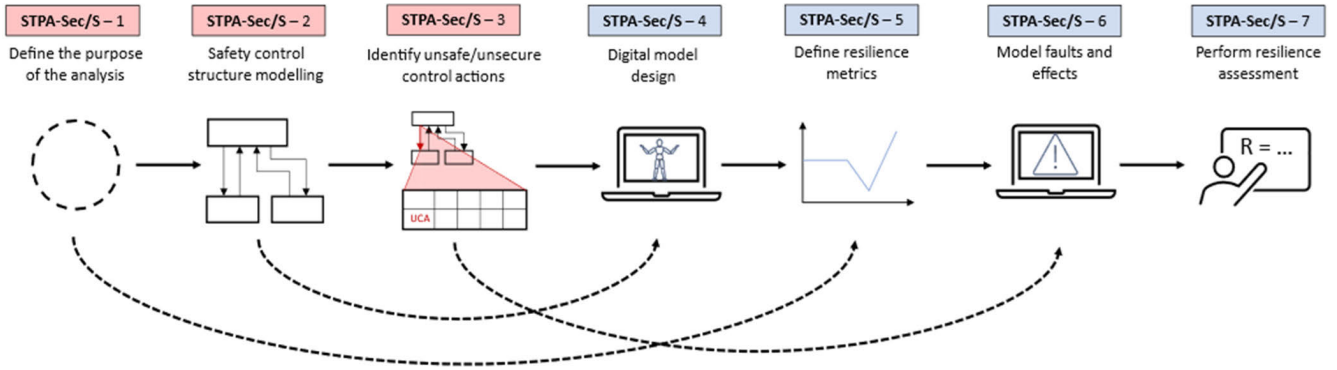

FIGURE 1. STPA-Sec/S methodological steps, as resulting from the integration of STPA-Sec technique (red boxes) with simulation-driven resilience assessment (blue boxes).

1) USING STAMP TO DEVELOP THE DIGITAL MODEL

The STAMP analysis helps developing a process model based upon the principles of systems theory. Through STAMP, it is possible to highlight all the system elements and their relationships in term of feedbacks they exchange, and control actions one imposes on another. This descriptive model can be used for other assessments, either qualitative as for STPA-Sec, or quantitative as for the STPA-Sec/S. For this purpose, the STAMP model is reproduced in a simulation environment to permit a cyber resilience assessment. The elements which require a digitalized counterpart are: (i) system entities and their state variables, (ii) controlled processes, (iii) sensors and feedbacks, (iv) controllers and related control algorithms, (v) actuators and control actions. Fig. 2 synthetizes the scope of this conversion.

The STAMP model identifies inputs and outputs for each controlled process. At first, it is necessary to highlight the entities to be processed inside the system, i.e., the elements which are transformed from being input to output. Notice how this identification (and the STAMP model itself) strictly depends upon the problem derived from purpose of the analysis. For example, in analyzing a water treatment plant, the entity to be modelled is clearly the water, which changes its state (i.e., is transformed) through the process. Accordingly, the entities to be reproduced in the simulation environment are the inputs and the outputs which are connected to the controlled processes in the STAMP model. The entities must be formalized mathematically through their characteristic dimensions, which are the state variables of the entities themselves. In the already mentioned example (i.e. a water treatment plant), state variables may comprehend (e.g.) water pressure, water flow rate, and water temperature. Thus, a system entity can be expressed as a vector $\overline{X}(t)$ of n state variables x:

$$\overline{X}(t) = (x_1, x_2, \ldots, x_n) \tag{1}$$

State variables are changed within the process since the entity is transformed being input (first) and output (then), and so on. It is then possible to define the entity $\overline{X}(t)$ in two

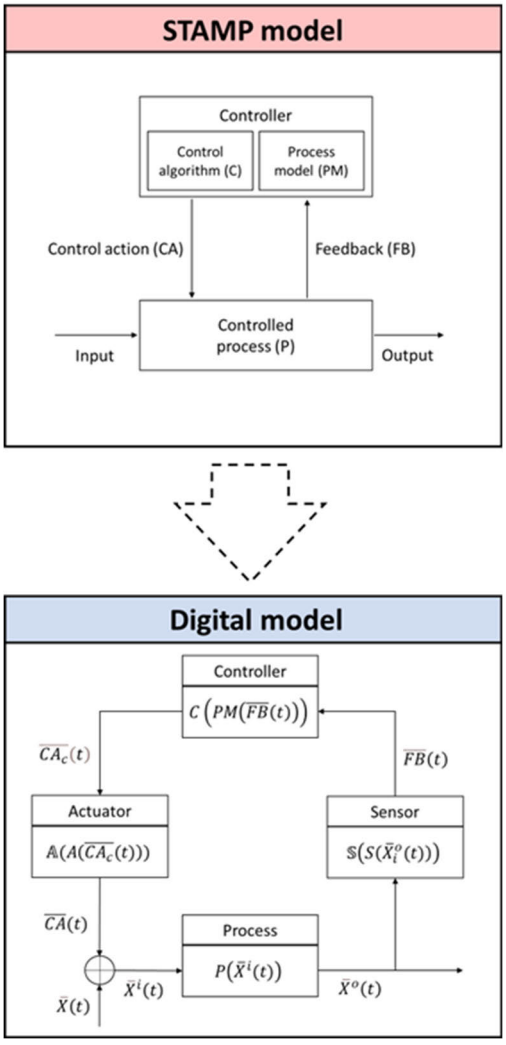

FIGURE 2. STAMP model conversion for the design of the digital model.

different stages of the process as:

$$\overline{X}^I(t) = \left(x_1^I, x_2^I, \ldots, x_n^I\right) \tag{2}$$

and:

$$\overline{X}^O(t) = \left(x_1^O, x_2^O, \ldots, x_n^O\right) \tag{3}$$

These latter representing two states of the entity, specifically before (i.e., $\overline{X}^I(t)$) and after (i.e., $\overline{X}^O(t)$) the transformation occurs (i.e., the process).

Processes are the elements of control, and they generate one or more outputs (both intermediate and final) throughout the system. The STAMP model provides an indication of processes to be modelled in the simulation environment. These latter equal the controlled processes from the STAMP model. A process is a part of the system in which the entity is transformed, and a change of its state variables is procured. Accordingly, the STAMP description of each controlled process, must be abstracted through a mathematical relation connecting $\overline{X}^I(t)$ to $\overline{X}^O(t)$:

$$\overline{X}^O(t) = P\left(\overline{X}^I(t)\right) \tag{4}$$

in which P is a transformation function that permits the change from a state to another.

Sensors are devices that responds to a signal or a stimulus and share the obtained information with the surrounding environment. Signals to be shared are the feedbacks highlighted in the STAMP representation. A sensor can be seen as a process with the aim to generate an output corresponding to a specific state variable measure. In the real world, this is done using well-known physical phenomena. In the modelling environment this can be done following a similar approach. Starting from process outputs, it is possible to obtain measures using equations that links physical quantities. The input of a sensors equals the output (or a part of it) of a related controlled process, so a relation can be written as:

$$\overline{FB}(t) = \mathbb{S}\left(S\left(\overline{X}^{O*}(t)\right)\right) \tag{5}$$

in which $\overline{FB}(t)$ represents the sensor output vectors (i.e., the feedback obtained from the sensing operations), $\overline{X}^{O*}(t)$ is the sensor input vector containing the entity state variable (or a group of state variables) after the controlled process has taken place, S is a function that allows computing the desired measures (e.g., a derivative to obtain acceleration from velocity), $\mathbb{S}$ is the characteristic function of the sensor device (i.e., the ability and the extent in which the device performs the measuring action effectively). Looking at a generic STAMP model, it is clear that not all process outputs are relevant in the calculus of the sensors output, i.e., not all the state variables are feedbacks. This consciousness justifies the need to use a $\overline{X}^{O*}(t)$ which is not necessarily equal to $\overline{X}^O(t)$. More specifically, some other observations can be made on $\overline{X}^O(t)$ vector. At first, not significant outputs, i.e., outputs that are not useful for modelling the sensor measure should be set to zero in $\overline{X}^{O*}(t)$. These state variables may be used to visualize system behavior in the final analysis, but they are not useful for model development. For some other state variables in $\overline{X}^O(t)$, no calculation or change might be needed, it happens when the modelled process gives as output the exact information that the sensor should read. In these cases, no transformation S is needed. Equation (5) can be so specialized relating each i-th element of the $\overline{FB}(t)$ vector to $\overline{X}^O(t)$. Accordingly:

$$\overline{FB}_i(t) = \begin{cases} 0 & for\ i : \overline{X}_i^O(t) is\ not\ in\ the\ feedback \\ S\left(\overline{X}_i^O(t)\right) & for\ i : \overline{X}_i^O(t) to\ be\ not\ transformed \\ \mathbb{S}\left(S\left(\overline{X}_i^O(t)\right)\right) & for\ i : \overline{X}_i^O(t)\ to\ be\ transformed \end{cases} \tag{6}$$

Feedbacks becomes inputs for the controllers, which aim to elaborate information and generate the control actions. In the STAMP model, a controller is defined by two parts: the process model, and the control algorithm. The controller's process model represents the knowledge the controller has on the system. It is updated through the feedbacks that the controller receives, and on a set of standard information about the system it is designed to have (i.e., how the process is meant to work). Based on the process model, the control algorithm produces the control action required to modify the controlled process. Control algorithms and their relationships with the process model (which relates to the feedback from the sensors) must be transposed from the STAMP model to the simulation environment. Accordingly, a controller can be formalized as a transformation C which produces a control action $\overline{CA_c}(t)$ based on a process model PM to be updated by the feedback $\overline{FB}(t)$:

$$\overline{CA_c}(t) = C\left(PM\left(\overline{FB}(t)\right)\right) \tag{7}$$

It is worth noticing that not all the state variables are modified by the control actions. Accordingly, the $\overline{CA_c}(t)$ vector must be arranged and filled with null values to be coherent with the entity state description dimension. To model the transformation C two approaches can be suggested. A first strategy may rely on the utilization of control theory. Considering the process model to be a dynamical system, a controller can be designed to make the system output follow a desired control signal. The system output will represent the control action to be imposed to the controlled process. The controller monitors this output comparing it with the reference input (i.e., the feedback) and adjusts its actions accordingly. An example may be continuous controller, such as proportional controllers, integral controllers, derivative controllers, and their combinations (e.g., PID controller). The second strategy may focus on the development of control algorithm through programming languages. The way a controller works can be seen as a response to a certain event involving the process. In this approach, the process model describes the occurrence or not of such events by computing state variables based on the feedback. The algorithm may trigger different control actions based upon the definition of a set of events and their occurrence, following certain logics.

The control action from the controller is not directly connected to the controlled process, but it is imposed to the process through an actuator. This latter is a device that works in reverse of a sensor, producing a signal or a stimulus from

an information obtained from the surroundings. Modelling actuators really depends on the accuracy the model is supposed to have. For simplicity, the output of the controller could be shaped as a signal that directly modify the process input. In case an in-depth study on actuators dynamics is needed, the modelling procedure follow a similar approach to the one used to model sensors. Physics laws permit to correlate the electronic signal entering the actuator to the transformed signal (may be physical or not) that will impact the process dynamics. Similar to sensors modelling, the correlation between actuators inputs and actuators outputs can be written in the form:

$$\overline{CA}(t) = \mathbb{A}(A\left(\overline{CA_c}(t)\right)) \tag{8}$$

In which $\overline{CA}(t)$ is the actuator's output vector representing the control action defined in the STAMP model. It contains values to modify the process' input vector $\overline{X}^I(t)$. $\overline{CA_c}(t)$ is the controller output vector (i.e., the control action prescribed by the controller), A is the function to correlate input and output and the physics transformation that occurs, and A is the characteristic function of the actuator device (i.e., the ability and the extent in which the device performs the control action effectively).

Finally, the modification of the controlled process input can be expressed by:

$$\overline{X}^I(t+1) = \overline{X}^I(t) + \overline{CA}(t) \tag{9}$$

being $\overline{X}^I(t+1)$ the state vector which describe the updated entity state at the entrance of the process.

2) USING STPA-SEC TO MODEL SYSTEM DISTURBANCES AND CAUSAL SCENARIOS

Based on the STAMP model, the STPA-Sec analysis provides the set of unsafe and unsecure control actions. These latter being critical to ensure both system safety and security. UCAs may derive from: (i) anomalies related to the controller, and (ii) anomalies related to the controlled process. The first case comprehends all those cases in which the controller processes a wrong control action due to a modification of its control algorithm. This can be done by altering the communication between two system parts, making them believe that they are directly communicating with each other. The attacker and its custom control algorithm insert themselves between these two system parts, acting as the real controller. The attacker who manages to prescribe its control logic into the system will disrupt the process taking control over it (e.g.), see man-in-the-middle attack strategies [57].

The controller may process a similar adverse outcome either if the control algorithm is modified by the attacker, or not. A modification to the input of the controller may lead to the generation of unsafe control actions. Inadequate feedback will update the process model wrongly, inducing to an unsafe/unsecure control even if the control algorithm is computing information correctly. Thus, anomalies related to the controlled process comprehends all the situation in which an UCA verifies due to wrong feedback to the controller. For cyber safety/security issues, this latter can rise from both unintentional events, and intentional events. The first set comprehend all the unpredictable external events which disrupt the controlled process feedback. An example may be heavy rainfall which may lead to communication blackout between sensors and controllers. This kind of event are strictly related to NaTech scenarios [58] and represents an inherent vulnerability of cyber-physical systems. The second set considers all situations where an adversary voluntarily forces a modification in the feedback, and consequently modifies the controller process model. For example, a cyber attacker may: hide the real sensor reading and inserting a wrong one inside the system, inducing wrong control actions, see (e.g.) false data injection attack strategies [59], [60]; or completely blocking communications throughout the IT system part with, for example, a jammer (e.g., denial of service attack strategies) [61]. Additionally, both anomalies related to controller, and anomalies related to the controlled process may occur simultaneously. The attacker may mask its adverse control actions by providing wrong feedbacks, remaining undetectable, see (e.g.) replay attack strategy [62].

Connecting these reasonings to the STAMP/STPA-Sec, the potential system failure scenarios can be derived from the unsafe control action prescribed by adverse control algorithm, but also from the inadequate feedback to the controller itself.

From the previous lines, it is clear how, in a cyber safety/security analysis, a relevant part of the simulation model is represented by communication systems. It is not necessarily true that a variable is perfectly moved from a component to another. For example, the feedback generated from the sensor reading may not be the same to be used as input by the controller. A perfect communication between two system elements is represented by an immediate and accurate connection that moves an information from the output of an element to the input of another. In this sense, the equations shown in the subsections above were built under the assumption of perfect communication. Communication has to be modelled taking into account two main concepts: differences in shape, i.e., exchange of information not following a certain communication protocol or specific rule, and differences in time, i.e., not immediate or not continuous exchange of information. Accordingly, based on (5) and (7) differences can be reproduced as:

$$\overline{FB}_{ATK}(t) = f_1(t) + f_2(t) \cdot \overline{FB}(t + f_3(t)) \tag{10}$$

$$\overline{CA}_{ATK}(t) = c_1(t) + c_2(t) \cdot \overline{CA}(t + c_3(t)) \tag{11}$$

being $f_1(t)$, $f_2(t)$, $c_1(t)$, and $c_2(t)$ elements of the time dependent vectors to describe the differences in shape of the feedback and the control action models respectively; and being $f_3(t)$, and $c_3(t)$ elements of the time dependent vectors to describe the differences in time for the feedback and the control action models respectively.

Even if these differences may be inherent in system functioning, a modification of them results in the occurrence of the UCAs. Accordingly, any not provided or wrongly provided CA (or FB) represents a difference in shape of the CA (or FB) model. Any CA (or FB) provided too late or too early, or for too long or too short, represents a difference in time of the CA (or FB) model. Modelling communication and their failures within the digital model allows reproducing different causal scenarios. For example, a not provided feedback at time t results in $f_1(t) = f_2(t) = 0$. A control action provided too late is reproduced through a $c_3(t) < 0$.

IV. CASE STUDY

The following section describes an application of the STPA-Sec/S methodology for a SeaWater Reverse Osmosis (SWRO) plant. At first, UCAs have been identified, underlining system criticalities. Interactions between system parts are depicted by the STAMP model. Later, a simulation model based on STAMP/STPA-Sec is used to quantify specific resilience metrics. The obtained results and their discussion have been provided.

This study focuses on the desalination process which is a water treatment process aiming at the removal of salts and minerals, and suspended solids from saline water to produce water suitable for human consumption. For demonstrative purposes, a SWRO plant represented the use case for STPA-Sec/S application. Specifically, seawater desalination is a separation process used to reduce the dissolved salt content of saline water to a usable level. All desalination processes involve three main water streams: the saline seawater feed stream, the low salinity produced water (i.e., permeate), and a very saline rejected concentrate (i.e., brine). The saline feedwater is drawn from oceanic or underground sources. It is then processed by the desalination process resulting in the two output streams (permeate, and brine). The permeate water is suitable for most domestic, industrial, and agricultural uses. The brine must be disposed generally by discharge into deep saline aquifers or surface waters with a higher salt content. The desalination process makes use of membranes to separate salt content to water. The most used are Reverse Osmosis (RO) membranes. A RO system performs four major processes (cf. Fig. 3):

- *Pre-treatment.* The incoming feedwater is pre-treated to be compatible with the membranes. Suspended solids of big dimension are removed by travelling screens (e.g., Travelling Screens (#4), Particle Settlements (#5), and Residual Treatment (#8)). Then the feedwater passes through a series of filters such as sand filters (#6), earth filters (#7), and cartridge filters (#10). Chemical processes to adjust the pH level of water occurs too.
- *Pressurization.* The pre-treated water has to reach from 5,5 MPa to 7 MPa for seawater desalination, high pressure pumps (#10) have to be used to achieve this result. The pumps raise the pressure to an operating value appropriate to guarantee the membrane passage based upon the salinity level of the feedwater.
- *Desalination.* The RO desalination process separates the pressurized feed stream water from the dissolved salts by let it flow through a water-permeable membrane (#11). The permeate is encouraged to flow through the membrane by the pressure differential created between the pressurized feedwater and the output stream, which is at atmospheric pressure. The permeable membranes inhibit the passage of dissolved salts while permitting the desalinated feed to pass through.
- *Post-treatment.* The product water of the membrane assembly usually requires pH adjustment and degasification before being stored (#14) to be transferred to a water distribution system (#16) to be used as drinkable water.

TABLE 1. Accident outcomes due to system hazards.

ID	Loss description	Category
L-01	Harmful or poisoned water supply	3rd party-human
L-02	Low quality water supply	3rd party-human
L-03	Reduction of permeate water	3rd party-human
L-04	Loss of reputation	Brand/relationship
L-05	Damage to equipment	Assets/equipment
L-06	Environmental contamination	Environmental
L-07	Financial loss	Legal/contractual
L-08	Water production stop	3rd party-human

A. STEP 1: DEFINE THE PURPOSE OF THE ANALYSIS

In accordance with the first step of the methodology, this section aims to define the system, and its operating scenario to be investigated [63]. These latter will be used to later create a control system model, and to identify the criticalities in the system itself. Accordingly:

- *Operating scenario.* The analysis will focus on the operating condition in which the SWRO plant is in steady-state production. The water quantities and water quality throughout the system are assumed to be constant during time if no disturbance applies.
- *System losses.* System losses organized by category are summarized in Table 1. These are the losses which turn out to have an essential value for stakeholders (the impacted stakeholders form each category).
- *System hazards.* The pumping system, the pre-treatment process, and the post-treatment process are the system operations highlighted to be critical for system losses. Table 2 contains the system hazards, in terms of their condition, behavior or period leading to the undesired event.

B. STEP 2: SAFETY CONTROL STRUCTURE MODELING

The second step prescribes the creation of the STAMP model. The model aim is to map the control procedures. Accordingly, at least two hierarchical levels are needed. A controller imposes constraints on the lower level, that same controller is

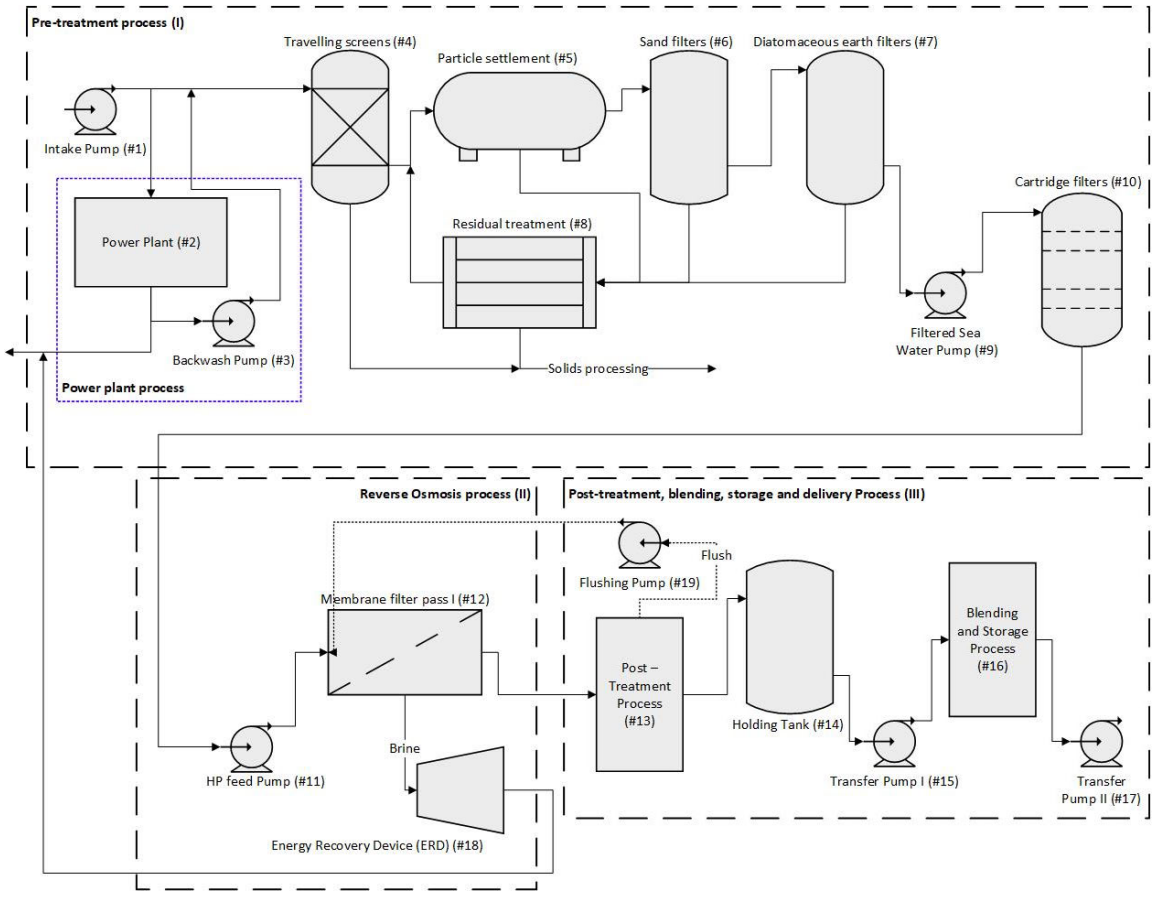

FIGURE 3. Seawater Reverse Osmosis plant (single stage) diagram.

TABLE 2. System hazards related to RO process.

ID	Hazards description	Related losses
H1	Pumping system imposes low pressure in the process	L-01, L-02, L-03, L-04, L-06, L-07, L-08
H2	Pumping system imposes high pressure in the process	L-03, L-05, L-07, L-08
H3	Not proper pre-treatment process	L-01, L-02, L-04,L-05, L-07, L-08
H4	Wrong chemical dosing in post-treatment stage	L-01, L-02, L-05, L-07

then controlled by higher level represented by the feedbacks and the control actions. In modelling the hierarchical control structure, a major focus must be dedicated on the control flow, since the inadequate control or feedbacks loops may result in system losses. The Safety Control Structure (SCS) for the SWRO plant has been created based on the system hazard and losses defined previously. A high-Level SCS is represented in Fig. 4 and comprehends:

- *SWRO plant central office* (green box in Fig. 4): comprehending the central utilities, the operation office; and the maintenance office. These latter being responsible for the internal organization, and for guaranteeing SWRO plant correct operational conditions.
- *SWRO plant auxiliary services* (purple box in Fig. 4): the RO plant is connected to a power plant to optimize the usage of resources and energy in the process. The SWRO plant deeply rely on electrical power to guarantee the correct operations, e.g., imposing an appropriate pressure on water is very energy consuming. Also, brine discharge offers a possibility for energy recovery. Accordingly, another controlled process is inserted in the high level SCS model, i.e., the power plant process.
- *SWRO plant process* (light blue box in Fig. 4): this section represents the core of the entire plant since it comprehends the components which allow the desalination process to physically take place. It comprehends: the SWRO plant operators crew (both operators and contractors), the SWRO plant central automated control system (i.e., SCADA systems), the automated control sub-systems for pre-treatment, the automated control sub-systems for reverse osmosis, the automated control sub-systems for post-treatment. The controlled processes are: the pre-treatment phase; the reverse osmosis phase, and the post-treatment – plus blending, storage and delivery – phase. The processes match with the ones depicted in Fig. 3.

Red boxes in Fig. 4 highlight the system parts that will be within the scope of the case study. Accordingly, a more granular SCS is proposed by isolating only the red boxes.

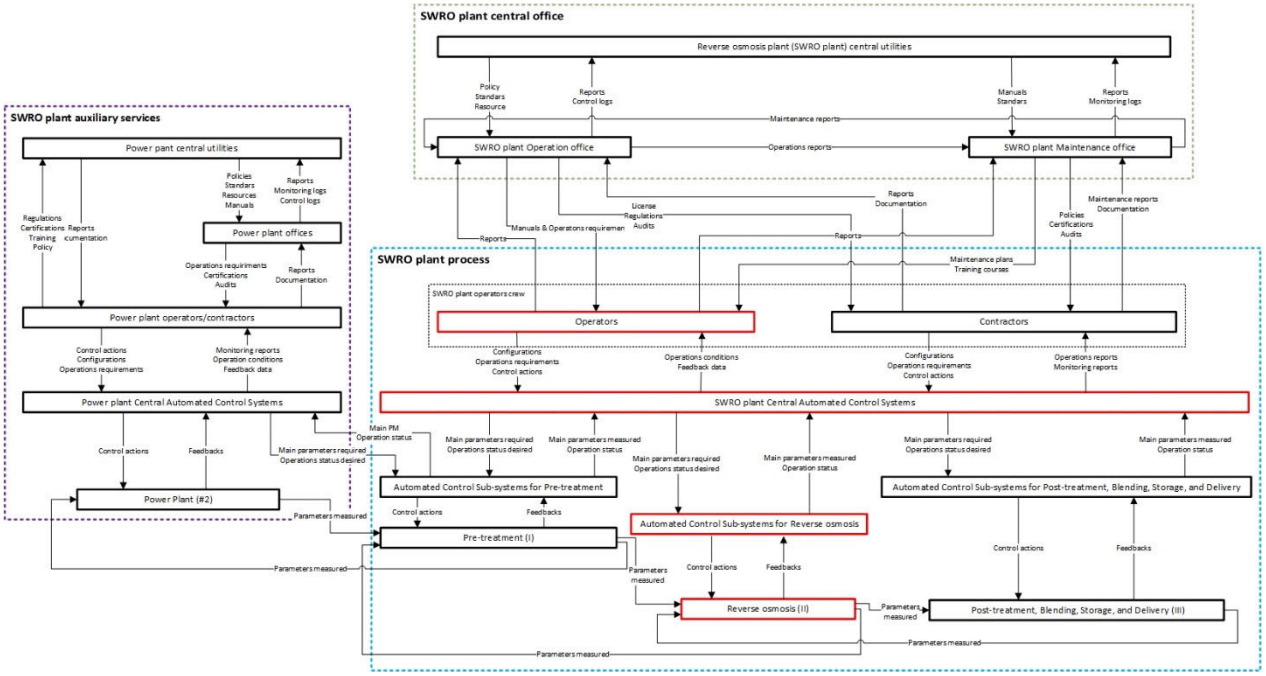

FIGURE 4. Seawater Reverse Osmosis process Safety Control Structure (red boxes depicts the system parts involved in the case study).

The fractal nature of STAMP allows exploiting controls at different levels of abstraction. This detailed SCS has been defined to highlight the control actions and feedback among: the High Pressure (HP) feed pump (#10), the Membrane filter pass (#11), the Energy recovery device (ERD) (#18), and their interactions. Fig. 4 shows the STAMP model isolating these components.

This excerpt has been further detailed in Fig. 6 in which further information concerning interactions and components has been inserted, too. At this level of detail, the controls have been explicated by means of two types of controllers. (i) Human controllers, represented by orange boxes in Fig. 6, that generate a control action on the Automated Controller and receives feedback regarding data from the controlled process. (ii) Automated controllers, represented by blue boxes in Fig. 6, which receive control action generated by the human controllers and forward this control in the process in light of their process model.

Additionally, the SWRO plant Central Automated Control System (light blue boxes cf. Fig. 6) shall guarantee the presence of a feedback loop on the process being controlled, as well as process operability in terms of correct actions.

C. STEP 3: IDENTIFY UNSAFE/UNSECURE CONTROL ACTIONS

The third step concerns the definition of the system-level hazards. These latter are identified by determining the system states or conditions that lead to a loss in worst-case operational and environmental circumstances. The hazards identified in the first part of the analysis can be linked with the UCAs. For the following evaluation the focus will be on the high pressure (HP) pump (#10) only, since this element play an important role in two out of four hazards: ''Pumping system impose low pressure in the process'', and ''Pumping system impose high pressure in the process''. Furthermore, the control actions ''RPM settings'', ''RPM value'' and ''RPM condition'' (red arrows, cf. Fig. 6) represent the control actions related to the HP feed pump.

Table 3 describes the causation (i.e., not provided, provided, timing or sequence, and duration) of each of those control actions.

The ''RPM condition'' control action (highlighted in italic font in Table 3) will be used to build and implement the resilience simulation analysis. This control action is responsible to handle the interactions between the Automated controller pump (#10) and the HP feed pump (#10). This is a crucial interface in the RO process since an inadequate action at this stage may lead the process to change the operating parameters significantly. Thus, this set of UCAs will be used to proceed with the cyber resilience analysis.

D. STEP 4: DIGITAL MODEL DESIGN

The digital model of a SWRO plant has been adapted from [54] reproducing the pressurization and desalination phases. The simulation model has been developed in the MATLAB/Simulink simulation environment. Simulink blocks have been re-arranged to be compliant with the STAMP descriptive model for the case study at hand. The model follows the principles of dynamic resilience modelling introduced in [64] as a dynamic approach to quantify resilience and resilience metrics under different stochastic conditions that can impact process performance.

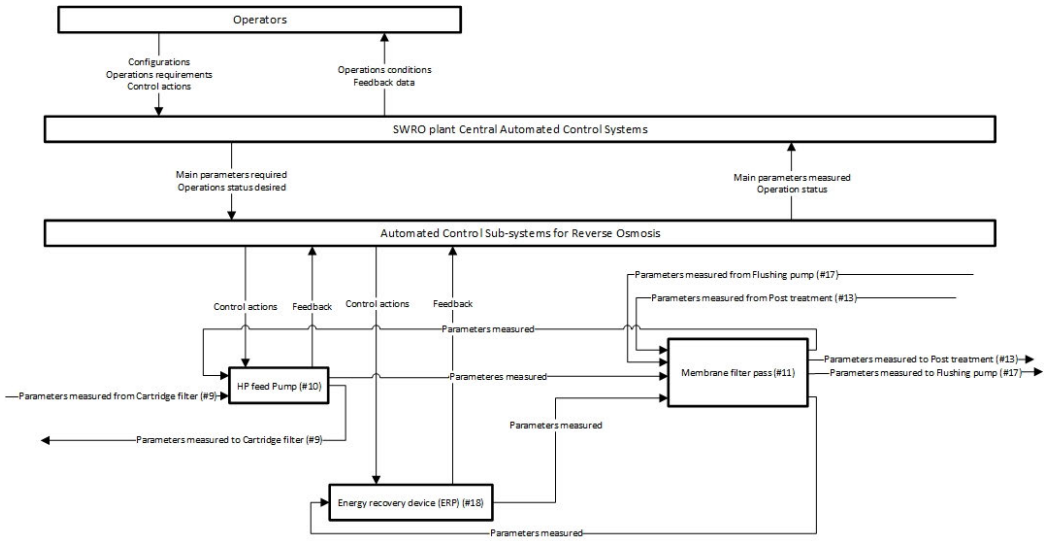

FIGURE 5. Safety Control Structure excerpt for system components under analysis.

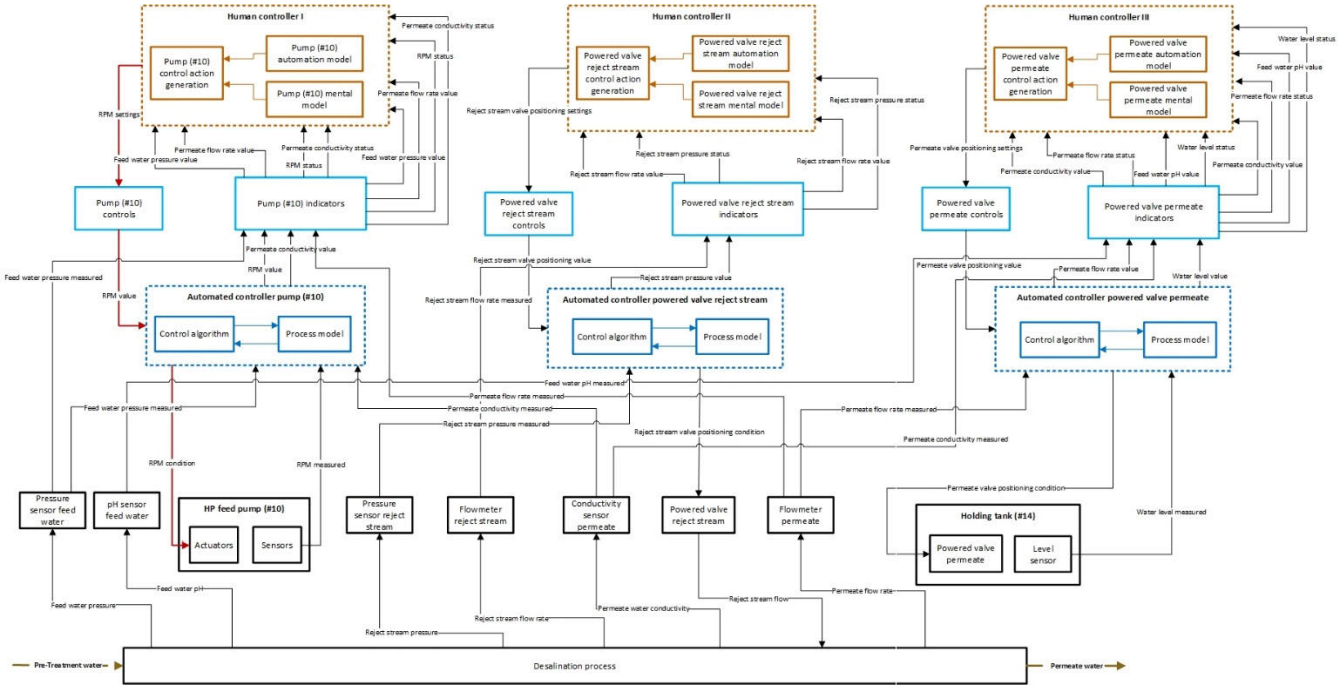

FIGURE 6. Detailed SCS for High Pressure pump (#10), Holding tank (#14), and Membrane filter pass (#11).

E. STEP 5: DEFINE RESILIENCE METRICS

The plant performance is evaluated based on the system losses identified in Table 1. Losses may concern the quality of the produced permeate (L-01, L-02, L-06), or the quantity of this latter (L-03, L-08). No emphasis is given to the plant loss of reputation (L-04) since it is mostly an indirect consequence of the quality/quantity losses. Damage to the equipment (L-05) and financial losses (L-07) are not quantified as well. Resilience metrics should allow the integration of system capacities and provide flexibility to capture system peculiarities [65]. Accordingly, two metrics are defined to evaluate the cyber resilient performance of the SWRO plant.

As simulation starts, the plant works in a defined steady state condition that is represented by a fixed flow rate of produced permeate water q_0. As permeate diminishes or increases, a coefficient depicting the variation related to production quantity can be calculated as follow:

$$M_1(t) = \frac{q_{(t)}}{q_0} \tag{12}$$

where $q_{(t)}$ is the permeate flow rate at simulation time t, and q_0 is the permeate flow rate at steady state working condition.

TABLE 3. UCAs description based on STAMP model.

Control action	Not provided	Provided	Timing or sequence	Duration
RPM settings	Human controller I does not set RPM parameters for the Pumps (10) controls.	Human controller I set RPM parameters inadequately	Human controller I set too early RPM settings for the Pump (10) controls	Not relevant for the analysis
			Human controller I set too late RPM settings for the Pump (10) controls	
RPM value	Pump (10) controls do not have communication with automated controller pump (10)	Pump (10) controls set RPM value inadequately	Pump (10) controls set too early RPM value with reference to feed water pressure measured	Not relevant for the analysis
	Pump (10) controls do not set RPM value		Pump (10) controls set too late RPM value with reference to feed water pressure measured	
RPM condition	*Automated controller pump (10) does not have communication with HP feed pump (10) actuators*	*Automated controller pump (10) set RPM condition wrong*	*Automated controller pump (10) set too early RPM condition with reference to permeate conductivity measured*	*Not relevant for the analysis*
	Automated controller pump (10) does not set RPM condition for the HP feed pump (10) actuators		*Automated controller pump (10) set too late RPM condition with reference to permeate conductivity measured*	

Similarly, the metric to evaluate loss in terms of quality of permeate water is defined as:

$$M_2(t) = \frac{C_{(t)}}{C_0} \tag{13}$$

where $C_{(t)}$ is the permeate conductivity at simulation time t, and C_0 is the permeate conductivity at steady state. The more water quality decreases (i.e., amount of salt in water increases and, subsequently, conductivity increases) the more $M_2(t)$ will increase. A value of the metric less than 1 does not necessarily depict a better working condition since SWRO is a very energy-expensive process. Basically, to provide a lower permeate conductivity the pressurization phase must consume more energy. $M_2(t) = 1$ is imposed also when the permeate water valve at the entrance of the storage tank is closed. In such configuration no water is entering the tank. The metric will so depict only the moments in which the quality of water will exceed the steady state requirements.

On the metrics time series, a measure of system cyber resilience for a specific performance is then given through the integral approach [66]:

$$R_i = \frac{\int_{t_0}^{t_1} M_{i_{(t)}} dt}{\int_{t_0}^{t_1} M^0_{i_{(t)}} dt} \tag{14}$$

where $M_{i_{(t)}}$ is the i-th metric time series, and $M^0_{i_{(t)}}$ is the steady state (i.e., not disrupted) performance curve. The resilience index R_i is equal to 1 if the two areas have the same extension, describing a perfect response with 100% resilience, it will decrease as long the difference between the two areas will increase, showing a less resilient response

F. STEP 6: MODEL FAULTS AND EFFECTS

Simulation scenarios to evaluate cyber-resilience are based on the development of an attacker model. It implements the capabilities of the opponent and can it be extended in order to reproduce different types of cyber attack strategies. It is assumed that the attacker can intercept any communication exchange throughout the model and so it can store, analyze, replay, alter and inject data. In this sense, cyber attacks are composed of two phases: a passive mode and an active mode. The passive phase aims to gain knowledge of the system, analyzing data without modifying information contained in them, but the adversary is already capable to provoke damages. Passive phase is not object of simulation since it is not interesting in evaluating system cyber-physical malfunctioning (that is the purpose of this study). Once an appropriate knowledge is obtained, the active mode begins. In this phase the attacker starts injecting data to take control of the system producing inadequate feedbacks, or/and hacked system controls that can generate different disruption scenarios.

Accordingly, based on the unsafe control action derived from the STPA-Sec analysis, two exemplary simulation scenarios are developed in line with the results in Table 3:

- *First simulation scenario.* The control action on HP pump provides wrong settings due to wrong feedback from sensors.
- *Second simulation scenario.* The attacker forces the control action on pump to provide wrong settings. Moreover, the feedback from sensors is masked by replaying previous measures.

1) FIRST ATTACK SCENARIO: SURGE ATTACK

The first scenario is reproduced through a surge cyber attack pattern [59]. It consists in a false data injection that aims to obtain maximum damage in the shortest time. False data

are inserted into the communication channel between sensors and controllers. In this way, controllers impose wrong control actions since corrupted sensor readings are provided to them. In this sense, a false pressure feedback may force plant shutdown since the RO desalination unit is designed to work under specific pressure range. Accordingly, the controller process model for the pump is able to consider this limit and to force pumps slowdown if the system reaches an alarming pressure. So, injecting a false pressure measure can result in: (i) controller forcing the pump to increase pressurization if the feedback reports a suboptimal pressure value (may lead to desalinator and pipes damages), (ii) controller forcing the pump to slow down with a subsequent loss on permeate quantity and quality, (iii) pump controller forcing the system to shut down if the feedback reports an alarming pressure (i.e., a pressure value that exceeds an imposed limit). The third situation is the one that most comply with the purpose of a surge attack to maximize damages in the shortest time. As long as the fake data (unacceptable pressure) are provided to the controller, pumps velocity will slow down. At first, the pump deceleration will cause a production decrease since a minor quantity of permeate will pass through filters. This happens since water enters the RO unit at lower pressure. In this first phase, good quality water is still produced, but in less quantity. At a second stage, water enters the RO unit with insufficient pressure. Some water will still passthrough filters, but its conductivity will be too high to consider the product acceptable. The downstream valve will be closed. The conductivity measure will suggest pumps controller to increase pump velocity but the pressure feedback still being unacceptable will maintain pump controller slowing down the pump. This phase already represents a production shutdown since no drinkable water is produced. If the adversary manages to transmit the false pressure measure for further time, the controller will completely stop the pump. Once this happens, the entire system is compromised causing a long disservice.

The scenario descripted above is modelled in the Simulink environment developing a custom block following the logic in the following pseudo-code:

Attack 1 Pseudo-Code (Surge Attack) – Feedback

Input:
P_f^m : measured feed flow pressure
P_{max} : maximum admittable pressure
t_{start} : cyber attack start time
t_{end} : cyber attack end time
t : current simulation time step
Output:
P_f^m : measured feed flow pressure

if $t_{start} < t\,(i) < t_{end}$ **then**
$P_f^m\,(i) = P_{max}$
else
$P_f^m\,(i) = P_f^m\,(i)$

In accordance with (10), and considering P_f^m to be the only value in the feedback vector $\overline{FB}$:

$$\overline{FB}_{ATK}\,(t) = f_1(t) = P_{max} \tag{15}$$

when the system is under attack, i.e., $t_{start} < t < t_{end}$.

2) SECOND ATTACK SCENARIO: REPLAY ATTACK

The second scenario is reproduced through a replay attack pattern [67]. It is structured in two phases that the adversary manages to perform simultaneously: (i) the attacker affects the feedback by replicating the last measure provided (which correspond to a normal operation condition), (ii) the attacker inserts adverse control actions to modify the system state. The attacker remains undetectable as long the feedback is replayed, since it will report a good working condition. A replay attack does not require a knowledge of system dynamic since once the access to sensors is obtained it simply continue replaying the previous measure. Replaying the feedback inhibits the controller, which will continue to analyze a good working condition and it completely loses the ability to know the actual system state. As a result, the controller process model is completely compromised. However, it still permit the calculation of good control outputs. So, also the connection between controllers and actuators is attacked. Wrong control actions are inserted on this communication branch and the system falls under the attacker control. Through the replay attack, the adversary can provoke different disruption. The simulation scenario is built with the purpose to create contaminated water, which is not highlighted by the plant sensors, and so it may reach the water distribution network. The conductivity feedback is the one replayed, making two controllers to be inhibited. The pump controller will consider the current velocity and the measured pressure to be good enough to produce good quality water, so it will not modify pump speed. The valve controller instead, will not close the path to the storage tank since the quality control will be passed. At the same time, the adverse control action is imposed on the pump. The attacker lowers pump velocity making the pressure fall, and increasing conductivity. As conductivity increases (as a result of a bad filtration process), flow rate will decrease, making the system producing fewer water at low quality. The product water will be considered pure enough to be supplied to population since the feedback will always provide a good conductivity measure. The scenario has been implemented into the Simulink simulation environment. Two custom blocks are developed, their logic is presented in the following pseudo-codes:

In accordance with (10) and (11), and considering C_p^m to be the only feedback, and v_{pump} to be the only control action in $\overline{FB}$ and $\overline{CA}$ respectively:

$$\overline{FB}_{ATK}\,(t) = \overline{FB}\,(t + f_3(t))\,, f_3\,(t) = -1 - (t - t_{start}) \tag{16}$$
$$\overline{CA}_{ATK}\,(t) = c_2\,(t) \cdot \overline{CA}\,(t)\,, c_2\,(t) = -1 \tag{17}$$

when the system is under attack, i.e., $t_{start} < t < t_{end}$.

Attack 2 Pseudo-Code (Replay Attack) – Feedback

Input:
C_p^m : measured permeate conductivity
t_{start} : cyber attack start time
t_{end} : cyber attack end time
t : current simulation time step
Output:
C_p^m : measured feed flow pressure

if $t_{start} < t\ (i) < t_{end}$ **then**
$C_p^m\ (i) = C_p^m\ (i-1)$
else
$C_p^m\ (i) = C_p^m\ (i)$

Attack 2 Pseudo-Code (Replay Attack) – Control Action

Input:
v_{pump} : desired pump velocity
v_{max}^{atk} : velocity target to be reached by the attacker
t_{start} : cyber attack start time
t_{end} : cyber attack end time
t : current simulation time step
Output:
v_{pump} : desired pump velocity

if $t_{start} < t\ (i) < t_{end}$ **and** $v_{pump} \neq v_{max}^{atk}$ **then**
$v_{pump}\ (i+1) = v_{pump}\ (i) \pm A \cdot i$
else
$v_{pump}(i) = v_{pump}(i)$

3) CYBER ATTACKS DURATION

The time between the moment in which the cyber attack starts, and the moment in which the system re-starts its normal working conditions gives a measure of the system recovery capacity. A probabilistic approach is used to model inherent variability in the process, starting from previously published works that helped electing reasonable time ranges, e.g., [18], [68]. The disruptions take place in a time range defined as:

$$\Delta T = t_{start} - t_{end} \tag{18}$$

where t_{start} identifies the moment in which the cyber-physical attack starts.

This means that all the tasks the adversary does to collect data, to gain knowledge and to enter the system are considered to be precedent to this moment (e.g., a phishing e-mail exchange to get sensitive information about plant functioning). On the other hand, t_{end} is the moment in which system functionality have been completely restored. This latter condition, does not imply that system restart to perform as it did before disruption occurs but that it is again the condition to do it. ΔT will give a measure of how quick the system is capable to recover. Concerning the system state after recovery, an as-good-as-before [69] logic is followed. This means that from t_{end}, the system is forced to improve performances since the pre-disruption state is reached again. Different ΔT will produce different metric patterns, Monte Carlo simulations are used to aggregate results. The simulation output will not be a deterministic value for the system cyber resilience but a set of them, related to the attack duration and their frequency. The number of simulation run is calculated conservatively through [70]. Accordingly, considering a 95% level of confidence the number of iterations is 218. Conservatively, 250 iterations are made.

G. STEP 7: PERFORM RESILIENCE ASSESSMENT

In this section presents the simulation outputs and the consequent cyber resilience assessment. The simulation scenarios are used to compute resilience metrics referred to both the proposed performance measures. Simulation have been conducted evaluating the system performances for a time frame of a week, without making any change in system normal condition behavior. This permits to underline effects of the proposed disruption scenarios on a medium/long term.

1) ATTACK SCENARIO 1: SIMULATION OUTCOMES

The first simulation refers to the surge attack. A wrong measure leads the system to shut down. This happens as long controllers identify a dangerous situation and cannot manage to resolve it by decreasing pump velocity, so pumps are stopped. Fig. 7 shows an exemplary pattern for this situation and consequent effects that such an attack has on the digital model outputs. In the specific case shown in Fig. 7, the attack is performed in the time range between $t_{start} = 60$ and $t_{end} = 600$.

Concerning contaminant, its value goes up, at first, since controller diminishes pump velocity to contain pressure. Contaminant increases because the lower pressure imposed on RO desalination unit implies a less pure permeate stream at its exit. The valve controller will close the valve once the conductivity measure will report an unacceptable value. Since no more water is entering the permeate water tank, the contaminant jumps to 0, i.e., no more contaminant is entering the tank (neither water too). At some point ($t = 600$ in this case) the threat is resolved, and the system starts working again as before. This means that the controller is aware of the real pressure measure, not the one imposed by the adversary, and it starts to accelerate pumps. This is not sufficient to repristinate system performance because water at not acceptable conductivity is still produced, and so the valve remains close. Once an acceptable conductivity is measured at desalinator exit, the valve is re-opened and, again, the controller regulates settings to return the system to its initial state. The loss on quantity represents the most critical performance since the cyber attack aims to stop water production for a certain time range. In a similar way to the contaminant profile, the volume loss profile has at first an increasing phase that is represented by pump deceleration. In facts, a lower pressure imposed at desalinator entrance implies a minor quantity of water passing through filters. Accordingly, a loss in production is verified. Once the conductivity sensor measures a

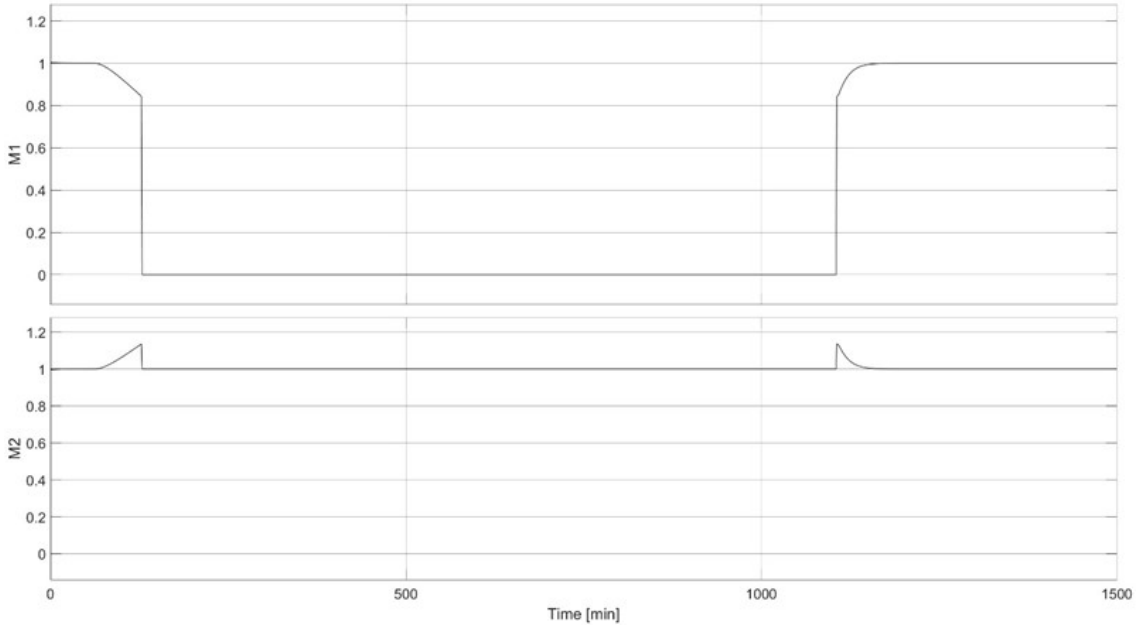

FIGURE 7. Exemplar M_1 and M_2 metrics' patterns for 1st attack scenario ($t_{start} = 60$, $t_{end} = 600$).

non-acceptable conductivity value, the valve is closed making the loss in production maximum (i.e., M_1 equal to 0). Again, when the system restarts its normal functioning, the initial performance is repristinated with a delay due to the pump restart.

2) ATTACK SCENARIO 2: SIMULATION OUTCOMES

The second simulation scenario reproduces the replay attack. It has the aim to both reduce production and insert low quality water in the water supply system. The conductivity measure is replayed in order to inhibit both the pump and the valve controllers. In this way valve controller will always drive water in the storage tank even if the adversary modify production by taking control on the system. Fig. 8 shows the effects of the attack on the Simulink model outputs. The disruption shown in Fig. 8 lasts from $t_{start} = 60$ to $t_{end} = 600$.

When the attack begins, the adversary starts to lower pumps velocity, this induces to both an increment on quantity of contaminant in produced water and a decrease in production (i.e., percentage loss increase). At some point, a new disrupted steady state is reached since the attacker managed to obtain the desired velocity decrease. The system works this way since the treat is resolved and the controllers regain awareness of system bad functioning. As long as the produced water is unacceptable from a quality point of view, the valve controller does not permit the water to enter the storage tank, this translates in a complete stop in production. The initial performance is reached again when the pumps reach an appropriate velocity, and the controllers sets process parameter to perform in an as-good-as-before state.

V. DISCUSSION ON RESULTS

System recovery phase is influenced by the random variables t_{start} and t_{end}. Accordingly, the computed resilience will be dependent from these parameters, too, with no possibility to assess it by running a single simulation. For this reason, multiple iterative simulations are performed to evaluate resilience as a function of random variables t_{start} and t_{end}.

Regarding the first simulation scenario, the results for the two resilience indicators are shown in Fig. 9. Table 4 reports numerical value for the indicators' distributions.

Decision makers from the plant management (or even regional water authorities) can utilize these results to identify worst case scenarios related to a specific control action failure. It is clear from Fig. 9 how this attack scenario has minimum impact on R_2. The loss on resilience is of the order of 10-4. In fact, the false pressure feedback injected to the pump controller leads it to turn off the pumps and stop the production, resulting in a minimum change of water quality inside the storage tank. The impact on water quantity is, on the contrary, huge. In the worst case a loss of more than 60% is registered. This may provoke major disservices to society if it is assumed that the storage tank is part of a distribution system. Confidence levels can be assigned to plant configuration, too. For example, in this configuration, more than the 65% on water production compliance is expected just in 25% of cases.

This value goes down to 55% if considering the median scenario (50% of cases).

Quite different results are obtained from the second simulation scenario (cf. Fig. 10, Table 5). In this case, the impact on water conductivity is clearly visible. In fact, the adversary aims to contaminate water being undetectable by masking the conductivity measure. The effect on water quality may be dangerous both from a societal and environmental point of view. SWRO plant usually are built in water critical regions in which freshwater is not available. So, desalinated water is used to feed people, but also for agriculture and

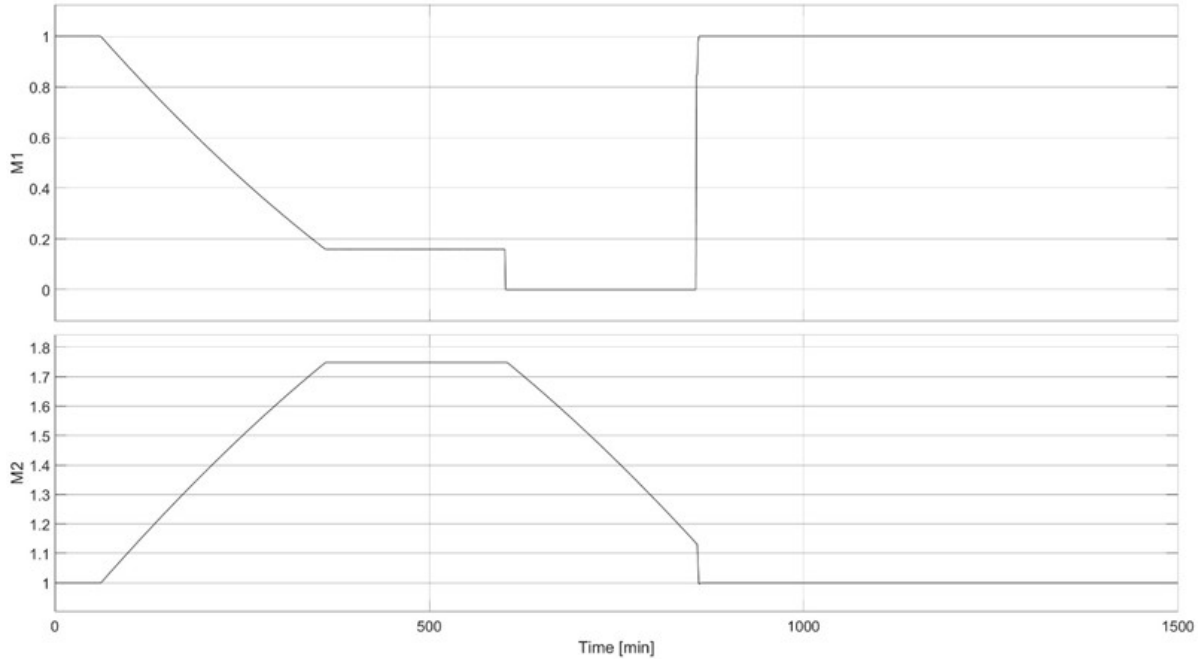

FIGURE 8. Exemplar M_1 and M_2 metrics' patterns for 2nd attack scenario ($t_{start} = 60$, $t_{end} = 600$).

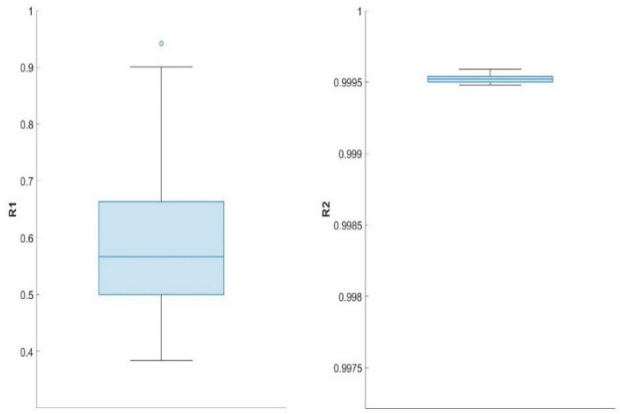

FIGURE 9. Boxplots for R_1 and R_2 cyber resilience indicators in 1st cyber attack scenario.

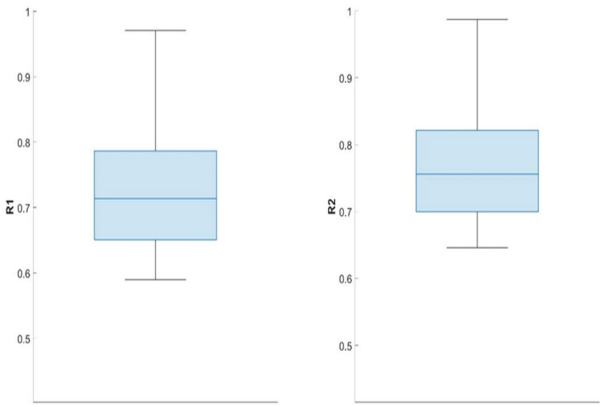

FIGURE 10. Boxplots for R_1 and R_2 cyber resilience indicators in 1st cyber attack scenario.

TABLE 4. Numerical data from box-plots in Fig. 9.

	R_1	R_2
Maximum value	0.9417	0.9996
75th percentile	0.6635	0.9995
Median value	0.5668	0.9995
25th percentile	0.4999	0.9995
Minimum value	0.3841	0.9995
Minimum value	0.3841	0.9995

TABLE 5. Numerical data from box-plots in Fig. 10.

	R_1	R_2
Maximum value	0.9706	0.9873
75th percentile	0.7868	0.8214
Median value	0.7136	0.7562
25th percentile	0.6506	0.7001
Minimum value	0.5901	0.6462
Minimum value	0.9706	0.9873

farming. Unacceptable quality standards may imply health consequences for people, and environmental contamination. In the simulated case study, there is almost no loss (2%) in the best-case condition, but just in the 25% of cases water quality compliance remain above 82%.

The median scenario prescribes a loss of 25% and, the worst case leads to a loss of more than 35%. Overall, the quantity loss depicted by R_1 is minor with respect to the previous scenario. The median value equals 71% compliance on production quantity, more than 65% of compliance with expected water production is obtained in 75% of cases, and in the worst case, a loss of 40% is depicted.

Simulation outputs also enable another type of analysis. Resilience values may be plotted with respect to cyber attack time to graphically define the plant's cyber resilience functions. For this sake, attack start time t_{start} and attack end

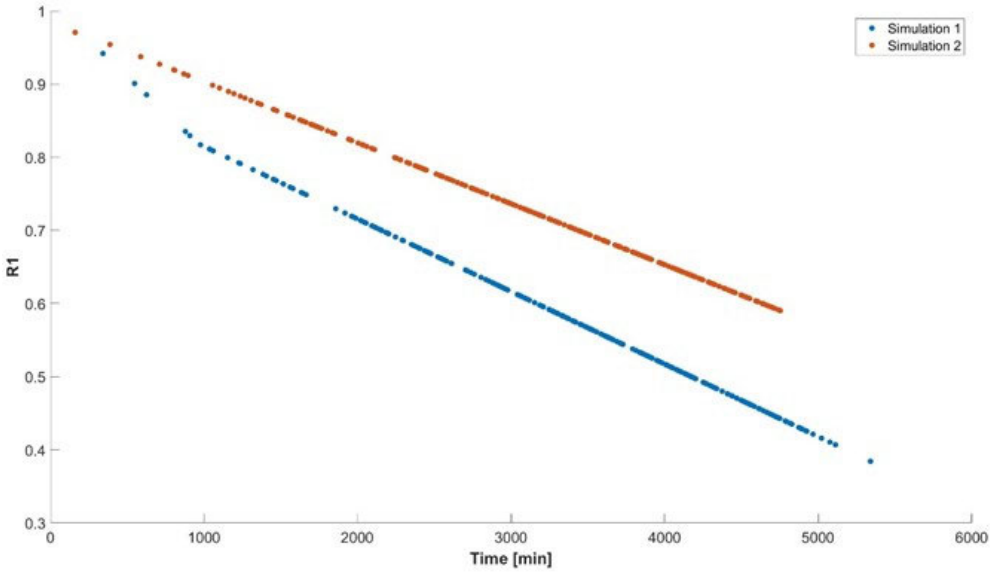

FIGURE 11. R_1 as a function of cyber attack duration for the two simulation scenarios.

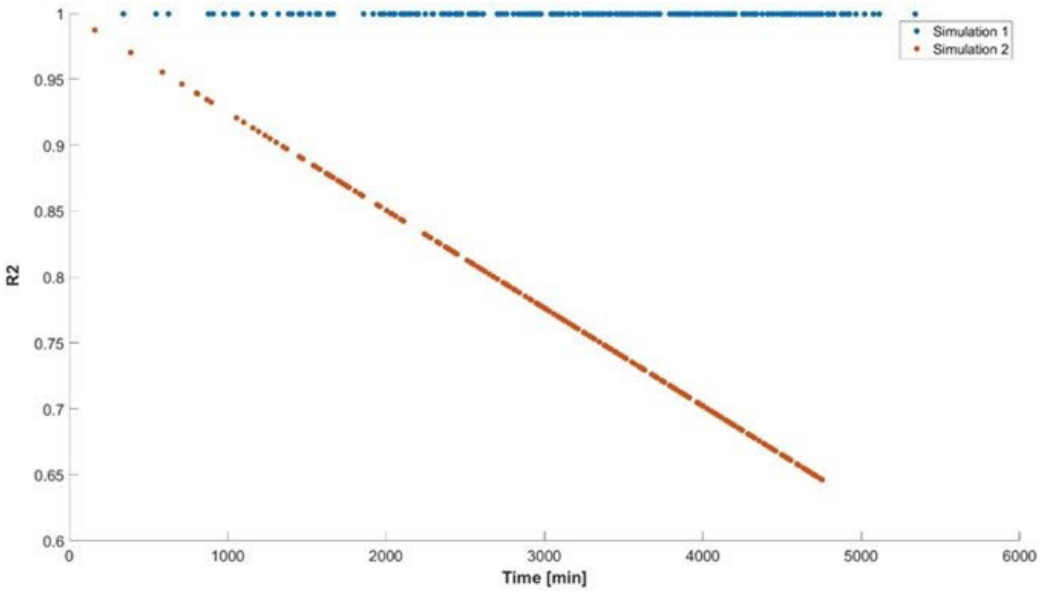

FIGURE 12. R_1 as a function of cyber attack duration for the two simulation scenarios.

time t_{end} have been aggregated following (18). Fig. 11 and Fig. 12 report respectively the R_1 and the R_2 resilience indicators over the two cyber attacks (different colors) durations. This type of results may guide decision making at arranging responses to cyber disruption by means of a desired level of service to maintain. For example, if a 90% performance on water quantity production has to be ensured, the cyber attacks may be contained and resolved within a range of 1000 minutes (almost 16 hours). Accordingly, this kind of reasoning can guide the business in quantifying the improvements to be made upon specific system parts to deal with system vulnerabilities highlighted in STPA-Sec analysis. This analysis can be used as well to compare multiple cyber attack scenarios to get the more critical in terms of impact on the water production. In this case, between the two, the second attack scenario resulted to be way more critical concerning the water quality, and quite less disruptive in terms of the quantity of water produced.

VI. CONCLUSION

STPA-Sec/S shows the possibility to quantify cyber resilience based on STAMP systems theoretic modelling. To do so, the STAMP model is converted into a simulation environment. This paper provides detailed guidance on the translation process of a STAMP safety control structure into its corresponding analytical entities needed for the simulation environment. The resulting STPA-Sec/S combines STPA-Sec with simulation models to develop and study a cyber-socio-technical system behavior under disruptions occurrence. The proposed methodology has been instantiated with a case study for a

seawater treatment facility. It is important to notice how the proposed industrial domain does not restrict the applicability of STPA-Sec/S in other industries. Studying cyber threats is not limited to the water supply sector since STPA-Sec/S may generate benefits for open research questions for e.g., nuclear plants [71], chemical plants [72], oil and gas industry [73], etc.

The obtained results demonstrate the feasibility of the proposed methodological solution to assess plant cyber resilience under two exemplary cyber attack scenarios. Successful cyber attacks against the systems that process potable water are significant since such failures may have public health and environmental impacts. Based on the industrial process of interest, various cyber attacks can be modeled considering different system's critical aspects. The proposed STPA-Sec/S analysis has no limit concerning cyber attack scenarios to be framed, permitting the cyber resilience assessment in multiple processes and operations settings.

Decision makers from water treatment facilities and water authorities may benefit of such methodology to take more secure decisions. Accordingly, such evaluations can be made at different stages of system lifecycle: (i) to guide the system engineering design and the development of process safety, integrating security countermeasures throughout system specifications; (ii) to address plant operation management in terms of adapting the process capabilities when a cyber attack occurs; (iii) to suggest operative countermeasures from a societal management perspective in terms of regulation, risk treatment plans, and risk assessment/mitigation procedures.

Despite this manuscript shows an assessment of cyber resilience for an industrial plant, its systemic perspective can be further reinforced. In this regard, future developments may include considering the fractal nature of resilience [74], exploring more dimensions, specifically:

- Micro-resilience, which refers to resilience of single system components, may be technical or a human.
- Meso-resilience, to consider resilient response of the whole organization.
- Macro-resilience, which is societal resilience, to extend impacts evaluation also considering society involvement and crowd behaviors.
- Cross-scale resilience, to consider not only societal impact but also environmental implications of system failures.

STPA-Sec/S fits each of these perspectives, taking advantage of the inner fractal nature of STAMP. In the case study, the problem to be analyzed leads to a focus on automated control structure, but a meso-resilience point of view might have considered also human controllers and organization above them to manage maintenance and recovery actions.

Similarly, no procedure to prioritize causal scenarios due to cyber attacks has been used. Future works may improve the proposed methodology by supporting a harmonized identification of the scenarios triggered by malicious cyber-physical attacks on plants [75]. Also, a prioritization step might be inserted as long it would be much time expensive running simulations for each scenario highlighted from STPA-Sec, specifically for highly complex systems.

If on one side, the cyber security problems have been recognized significant for critical infrastructures [76], on the other, the innovations related to the use of CPS and security-related technologies in process safety are still not explored deeply [77]. Overall, the results of this study contribute to the industrial engineering, setting a staging area to evolve quantitative cyber resilience assessment for process plants, putting into operational terms systems theory for process engineering design and practice.

REFERENCES

[1] T. Stock, M. Obenaus, S. Kunz, and H. Kohl, "Industry 4.0 as enabler for a sustainable development: A qualitative assessment of its ecological and social potential," *Process Saf. Environ. Protection*, vol. 118, pp. 254–267, Aug. 2018, doi: 10.1016/j.psep.2018.06.026.

[2] S. Dekker, *The Field Guide to Understanding 'Human Error'*. Farnham, U.K.: Ashgate Publishing Company, 2014.

[3] E. Trist, "The relations of social and technical systems in coal-mining," British Psychological Society, London, U.K., 1950, pp. 1–67.

[4] R. Patriarca, A. Falegnami, F. Costantino, G. Di Gravio, A. De Nicola, and M. L. Villani, "WAx: An integrated conceptual framework for the analysis of cyber-socio-technical systems," *Saf. Sci.*, vol. 136, Apr. 2021, Art. no. 105142, doi: 10.1016/j.ssci.2020.105142.

[5] N. Bugalia, Y. Maemura, and K. Ozawa, "Organizational and institutional factors affecting high-speed rail safety in Japan," *Saf. Sci.*, vol. 128, Aug. 2020, Art. no. 104762, doi: 10.1016/j.ssci.2020.104762.

[6] P. Underwood and P. Waterson, "Systems thinking, the Swiss Cheese model and accident analysis: A comparative systemic analysis of the Grayrigg train derailment using the ATSB, AcciMap and STAMP models," *Accident Anal. Prevention*, vol. 68, no. 1, pp. 75–94, Jul. 2014, doi: 10.1016/j.aap.2013.07.027.

[7] N. Paltrinieri, S. Bonvicini, G. Spadoni, and V. Cozzani, "Cost-benefit analysis of passive fire protections in road LPG transportation: Cost-benefit analysis of passive fire protections," *Risk Anal.*, vol. 32, no. 2, pp. 200–219, Feb. 2012, doi: 10.1111/j.1539-6924.2011.01654.x.

[8] P. V. R. de Carvalho, "The use of functional resonance analysis method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience," *Rel. Eng. Syst. Saf.*, vol. 96, no. 11, pp. 1482–1498, Nov. 2011, doi: 10.1016/j.ress.2011.05.009.

[9] M. Span, L. O. Mailloux, R. F. Mills, and W. Young, "Conceptual systems security requirements analysis: Aerial refueling case study," *IEEE Access*, vol. 6, pp. 46668–46682, 2018, doi: 10.1109/ACCESS.2018.2865736.

[10] N. Leveson, "A new accident model for engineering safer systems," *Saf. Sci.*, vol. 42, no. 4, pp. 237–270, 2004, doi: 10.1016/S0925-7535(03)00047-X.

[11] W. Li, L. Zhang, and W. Liang, "An accident causation analysis and taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives," *Saf. Sci.*, vol. 92, pp. 94–103, Feb. 2017, doi: 10.1016/j.ssci.2016.10.001.

[12] N. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA, USA: MIT Press, 2012, doi: 10.5860/choice.49-6305.

[13] R. Patriarca, M. Chatzimichailidou, N. Karanikas, and G. Di Gravio, "The past and present of system-theoretic accident model and processes (STAMP) and its associated techniques: A scoping review," *Saf. Sci.*, vol. 146, Feb. 2022, Art. no. 105566, doi: 10.1016/j.ssci.2021.105566.

[14] J. Rasmussen, "Risk management in a dynamic society: A modeling problem," *Saf. Sci.*, vol. 27, no. 3, pp. 183–213, Mar. 1997.

[15] N. Leveson, *STPA Handbook*. Cambridge, MA, USA: MIT Press, 2018.

[16] F. Björck, M. Henkel, J. Stirna, and J. Zdravkovic, "Cyber resilience—Fundamentals for a definition," *Adv. Intell. Syst. Comput.*, vol. 353, pp. 311–316, Jan. 2015, doi: 10.1007/978-3-319-16486-1_31.

[17] C. Cimpanu. (2021). *Two More Cyber-Attacks Hit Israel's Water System*. ZDNet. Accessed: Jun. 5, 2022. [Online]. Available: https://www.zdnet.com/article/two-more-cyber-attacks-hit-israels-water-system/

[18] A. Hassanzadeh, A. Rasekh, S. Galelli, M. Aghashahi, R. Taormina, A. Ostfeld, and M. K. Bank, "A review of cybersecurity incidents in the water sector," *J. Environ. Eng., United States*, vol. 146, no. 5, 2020, Art. no. 03120003, doi: 10.1061/(ASCE)EE.1943-7870.0001686.

[19] X. Liu, Z. Zhao, H. Li, C. Liu, and S. Wang, "Defect prediction of radar system software based on bug repositories and behavior models," *Int. J. Performability Eng.*, vol. 16, no. 2, pp. 284–296, 2020, doi: 10.23940/ijpe.20.02.p11.284296.

[20] M. Lower, J. Magott, and J. Skorupski, "A system-theoretic accident model and process with human factors analysis and classification system taxonomy," *Saf. Sci.*, vol. 110, pp. 393–410, Dec. 2018, doi: 10.1016/j.ssci.2018.04.015.

[21] Y. Lu, S.-G. Zhang, P. Tang, and L. Gong, "STAMP-based safety control approach for flight testing of a low-cost unmanned subscale blended-wing-body demonstrator," *Saf. Sci.*, vol. 74, pp. 102–113, Apr. 2015, doi: 10.1016/j.ssci.2014.12.005.

[22] G. C. Lordos, S. E. Summers, J. A. Hoffman, and O. L. De Weck, "Human–machine interactions in Apollo and lessons learned for living off the land on Mars," in *Proc. IEEE Aerosp. Conf.*, Mar. 2019, pp. 1–17, doi: 10.1109/AERO.2019.8741618.

[23] A. Kothakonda and J. Robertson, "CAST analysis of the international space station EVA 23 suit water intrusion mishap," in *Proc. Int. Astron. Congr.*, Oct. 2018, pp. 1–8.

[24] A. Al-Barnawi, Y. He, L. A. Maglaras, and H. Janicke, "Electronic medical records and risk management in hospitals of Saudi Arabia," *Informat. Health Social Care*, vol. 44, no. 2, pp. 189–203, Apr. 2019, doi: 10.1080/17538157.2018.1434181.

[25] R. Hosse, S. Sikatzki, E. Schnieder, and N. Bandelow, "Increasing systems-safety by meliorating policy-processes under conditions of ambiguity analyzing interdisciplinary ascendancies of the German traffic system by using cybernetic hazard analyzing methodologies," in *Proc. 3rd Int. Multi-Conf. Complex., Informat. Cybern.*, 2012, pp. 374–379.

[26] G. J. M. Read, A. Naweed, and P. M. Salmon, "Complexity on the rails: A systems-based approach to understanding safety management in rail transport," *Rel. Eng. Syst. Saf.*, vol. 188, pp. 352–365, Aug. 2019, doi: 10.1016/j.ress.2019.03.038.

[27] S. H. Lee, S.-M. Shin, J. S. Hwang, and J. Park, "Operational vulnerability identification procedure for nuclear facilities using STAMP/STPA," *IEEE Access*, vol. 8, pp. 166034–166046, 2020, doi: 10.1109/ACCESS.2020.3021741.

[28] A. Yousefi and M. R. Hernandez, "Using a system theory based method (STAMP) for hazard analysis in process industry," *J. Loss Prevention Process Industries*, vol. 61, pp. 305–324, Sep. 2019, doi: 10.1016/j.jlp.2019.06.014.

[29] J. R. Laracy and N. G. Leveson, "Apply STAMP to critical infrastructure protection," in *Proc. IEEE Conf. Technol. Homeland Secur.*, May 2007, pp. 215–220, doi: 10.1109/THS.2007.370048.

[30] W. Young and N. Leveson, "Systems thinking for safety and security," in *Proc. 29th Annu. Comput. Secur. Appl. Conf.*, Dec. 2013, pp. 1–8, doi: 10.1145/2523649.2530277.

[31] J. M. Sayers, B. E. Feighery, and M. T. Span, "A STPA-Sec case study: Eliciting early security requirements for a small unmanned aerial system," in *Proc. IEEE Syst. Secur. Symp. (SSS)*, Jul. 2020, pp. 1–7, doi: 10.1109/SSS47320.2020.9197728.

[32] L. O. Mailloux, M. Span, R. F. Mills, and W. Young, "A top down approach for eliciting systems security requirements for a notional autonomous space system," in *Proc. IEEE Int. Syst. Conf. (SysCon)*, Apr. 2019, pp. 1–7, doi: 10.1109/SYSCON.2019.8836929.

[33] P. Beaumont and S. Wolthusen, "Micro-grid control security analysis: Analysis of current and emerging vulnerabilities," in *Critical Infrastructure Security and Resilience*, Jan. 2019, pp. 159–184, doi: 10.1007/978-3-030-00024-0_9.

[34] S. Sharma, A. Flores, C. Hobbs, J. Stafford, and S. Fischmeister, "Safety and security analysis of AEB for L4 autonomous vehicle using STPA," *OpenAccess Ser. Informat.*, vol. 68, no. 5, pp. 1–5, 2019, doi: 10.4230/OASIcs.ASD.2019.5.

[35] J. Ge, Y. Zhang, K. Xu, J. Li, X. Yao, C. Wu, S. Li, F. Yan, J. Zhang, and Q. Xu, "A new accident causation theory based on systems thinking and its systemic accident analysis method of work systems," *Process Saf. Environ. Protection*, vol. 158, pp. 644–660, Feb. 2022, doi: 10.1016/j.psep.2021.12.036.

[36] A. Yousefi and M. R. Hernandez, "A novel methodology to measure safety level of a process plant using a system theory based method (STAMP)," *Process Saf. Environ. Protection*, vol. 136, pp. 296–309, Apr. 2020, doi: 10.1016/j.psep.2020.01.035.

[37] F. G. R. D. Souza, C. M. Hirata, and S. Nadjm-Tehrani, "Synthesis of a controller algorithm for safety-critical systems," *IEEE Access*, vol. 10, pp. 76351–76375, 2022, doi: 10.1109/ACCESS.2022.3192436.

[38] A. Abdulkhaleq and S. Wagner, "Integrated safety analysis using systems-theoretic process analysis and software model checking," in *Proc. Int. Conf. Comput. Saf., Rel., Secur.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9337, 2015, pp. 121–134, doi: 10.1007/978-3-319-24255-2_10.

[39] M. Tsuji, T. Takai, K. Kakimoto, N. Ishihama, M. Katahira, and H. Iida, "Prioritizing scenarios based on STAMP/STPA using statistical model checking," in *Proc. IEEE Int. Conf. Softw. Test., Verification Validation Workshops (ICSTW)*, Oct. 2020, pp. 124–132, doi: 10.1109/ICSTW50294.2020.00032.

[40] A. L. Dakwat and E. Villani, "System safety assessment based on STPA and model checking," *Saf. Sci.*, vol. 109, pp. 130–143, Nov. 2018, doi: 10.1016/j.ssci.2018.05.009.

[41] A. Thapaliya and G. Kwon, *Realization of Combined Systemic Safety Analysis of Adverse Train Control System Using Model Checking*, vol. 542. Singapore: Springer, 2019, doi: 10.1007/978-981-13-3648-5_49.

[42] D. Suo, J. An, and J. Zhu, "A new approach to improve safety of reconfiguration in integrated modular avionics," in *Proc. IEEE/AIAA 30th Digit. Avionics Syst. Conf.*, Oct. 2011, pp. 1–4, doi: 10.1109/DASC.2011.6095970.

[43] J. Zhang, H. Kim, Y. Liu, and M. A. Lundteigen, "Combining system-theoretic process analysis and availability assessment: A subsea case study," *Proc. Inst. Mech. Eng., O, J. Risk Rel.*, vol. 233, no. 4, pp. 520–536, Aug. 2019, doi: 10.1177/1748006X18822224.

[44] H. Sun, H. Wang, M. Yang, and G. Reniers, "A STAMP-based approach to quantitative resilience assessment of chemical process systems," *Rel. Eng. Syst. Saf.*, vol. 222, Jun. 2022, Art. no. 108397, doi: 10.1016/j.ress.2022.108397.

[45] C. W. Lee and S. Madnick, "Cybersafety approach to cybersecurity analysis and mitigation for mobility-as-a-service and Internet of Vehicles," *Electronics*, vol. 10, no. 10, pp. 1–22, 2021, doi: 10.3390/electronics10101220.

[46] M. T. Span, L. O. Mailloux, M. R. Grimaila, and W. B. Young, "A systems security approach for requirements analysis of complex cyber-physical systems," in *Proc. Int. Conf. Cyber Secur. Protection Digit. Services*, Jun. 2018, pp. 1–8, doi: 10.1109/CyberSecPODS.2018.8560682.

[47] J. Yu, Y. Jinghua, F. Luo, and S. Wagner, "Data-flow-based adaption of the system-theoretic process analysis for security (STPA-Sec)," *PeerJ Comput. Sci.*, vol. 7, pp. 1–21, Jan. 2021, doi: 10.7717/peerj-cs.362.

[48] S. S. Shapiro, "Privacy risk analysis based on system control structures: Adapting system-theoretic process analysis for privacy engineering," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2016, pp. 17–24, doi: 10.1109/SPW.2016.15.

[49] D. P. Pereira, C. Hirata, and S. Nadjm-Tehrani, "A STAMP-based ontology approach to support safety and security analyses," *J. Inf. Secur. Appl.*, vol. 47, pp. 302–319, Aug. 2019, doi: 10.1016/j.jisa.2019.05.014.

[50] K. A. Klise, M. Bynum, D. Moriarty, and R. Murray, "A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study," *Environ. Model. Softw.*, vol. 95, pp. 420–431, Sep. 2017, doi: 10.1016/j.envsoft.2017.06.022.

[51] N. Rashid, J. Wan, G. Quiros, A. Canedo, and M. A. A. Faruque, "Modeling and simulation of cyberattacks for resilient cyber-physical systems," in *Proc. 13th IEEE Conf. Autom. Sci. Eng. (CASE)*, Aug. 2017, pp. 988–993, doi: 10.1109/COASE.2017.8256231.

[52] J. Wan, A. Canedo, and M. A. Al Faruque, "Security-aware functional modeling of cyber-physical systems," in *Proc. IEEE 20th Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2015, pp. 1–4, doi: 10.1109/ETFA.2015.7301644.

[53] I. Tomic, M. Breza, and J. A. McCann, "Jamming-resilient control and communication framework for cyber physical systems," in *Proc. Living Internet Things*, 2019, pp. 1–6, doi: 10.1049/cp.2019.0132.

[54] F. Simone and R. Patriarca, "A simulation-driven cyber resilience assessment for water treatment plants," in *Proc. 32nd Eur. Saf. Rel. Conf.*, 2022, pp. 1–15.

[55] A. M. Law, "How to build valid and credible simulation models," in *Proc. Winter Simulation Conf. (WSC)*, Dec. 2019, pp. 1402–1414, doi: 10.1109/WSC40007.2019.9004789.

[56] K. H. Blanchard and S. Johnson. (1983). *The One Minute Manager*. Berkley Books. [Online]. Available: https://books.google.it/books?id=sxA9KMIJGZgC
[57] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Syst.*, vol. 35, no. 1, pp. 82–92, Feb. 2015, doi: 10.1109/MCS.2014.2364723.
[58] F. Ricci, V. C. Moreno, and V. Cozzani, "A comprehensive analysis of the occurrence of Natech events in the process industry," *Process Saf. Environ. Protection*, vol. 147, pp. 703–713, Mar. 2021, doi: 10.1016/j.psep.2020.12.031.
[59] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Secur.*, Mar. 2011, pp. 355–366, doi: 10.1145/1966913.1966959.
[60] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–33, May 2011, doi: 10.1145/1952982.1952995.
[61] X. Cai, K. Shi, K. She, S. Zhong, Y. Soh, and Y. Yu, "Performance error estimation and elastic integral event triggering mechanism design for T-S fuzzy networked control system under DoS attacks," *IEEE Trans. Fuzzy Syst.*, early access, Aug. 18, 2022, doi: 10.1109/TFUZZ.2022.3199817.
[62] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, Jul. 2014, doi: 10.1109/TCST.2013.2280899.
[63] Q.-Y. Zhang, L.-S. Liu, and Z.-J. Liu, "Application of safety and reliability analysis in wastewater reclamation system," *Process Saf. Environ. Protection*, vol. 146, pp. 338–349, Feb. 2021, doi: 10.1016/j.psep.2020.09.010.
[64] M. Taleb-Berrouane and F. Khan, "Dynamic resilience modelling of process systems," *Chem. Eng. Trans.*, vol. 77, pp. 313–318, Jan. 2019, doi: 10.3303/CET1977053.
[65] R. Yarveisy, C. Gao, and F. Khan, "A simple yet robust resilience assessment metrics," *Rel. Eng. Syst. Saf.*, vol. 197, May 2020, Art. no. 106810, doi: 10.1016/j.ress.2020.106810.
[66] C. Poulin and M. B. Kane, "Infrastructure resilience curves: Performance measures and summary metrics," *Rel. Eng. Syst. Saf.*, vol. 216, Dec. 2021, Art. no. 107926, doi: 10.1016/j.ress.2021.107926.
[67] D. Ding, Q.-L. Han, Y. Xiang, C. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, Jan. 2018, doi: 10.1016/j.neucom.2017.10.009.
[68] N. Tuptuk, P. Hazell, J. Watson, and S. Hailes, "A systematic review of the state of cyber-security in water systems," *Water*, vol. 13, no. 1, p. 81, Jan. 2021, doi: 10.3390/w13010081.
[69] B. M. Ayyub, "Practical resilience metrics for planning, design, and decision making," *ASCE-ASME J. Uncertainty Eng. Syst., A, Civil Eng.*, vol. 1, no. 3, Sep. 2015, Art. no. 04015008, doi: 10.1061/AJRUA6.0000826.
[70] G. Di Gravio, M. Mancini, R. Patriarca, and F. Costantino, "Overall safety performance of air traffic management system: Forecasting and monitoring," *Saf. Sci.*, vol. 72, pp. 351–362, Feb. 2015, doi: 10.1016/j.ssci.2014.10.003.
[71] S. Kim, G. Heo, E. Zio, J. Shin, and J.-G. Song, "Cyber attack taxonomy for digital environment in nuclear power plants," *Nucl. Eng. Technol.*, vol. 52, no. 5, pp. 995–1001, May 2020, doi: 10.1016/j.net.2019.11.001.
[72] P. R. Dunaka and B. Mcmillin, "Cyber-physical security of a chemical plant," in *Proc. IEEE 18th Int. Symp. High Assurance Syst. Eng. (HASE)*, Jan. 2017, pp. 33–40, doi: 10.1109/HASE.2017.23.
[73] A. Srivastava and J. P. Gupta, "New methodologies for security risk assessment of oil and gas industry," *Process Saf. Environ. Protection*, vol. 88, no. 6, pp. 407–412, Nov. 2010, doi: 10.1016/j.psep.2010.06.004.
[74] J. Bergström and S. W. A. Dekker, "Bridging the macro and the micro by considering the Meso: Reflections on the fractal nature of resilience," *Ecol. Soc.*, vol. 19, no. 4, pp. 1–9, 2014, doi: 10.5751/ES-06956-190422.
[75] F. Yan, J. Ma, M. Li, R. Niu, and T. Tang, "An automated accident causal scenario identification method for fully automatic operation system based on STPA," *IEEE Access*, vol. 9, pp. 11051–11064, 2021, doi: 10.1109/ACCESS.2021.3050472.
[76] Y. F. Khalil, "A novel probabilistically timed dynamic model for physical security attack scenarios on critical infrastructures," *Process Saf. Environ. Protection*, vol. 102, pp. 473–484, Jul. 2016, doi: 10.1016/j.psep.2016.05.001.
[77] J. A. Gobbo, C. M. Busso, S. C. O. Gobbo, and H. Carreão, "Making the links among environmental protection, process safety, and industry 4.0," *Process Saf. Environ. Protection*, vol. 117, pp. 372–382, Jul. 2018, doi: 10.1016/j.psep.2018.05.017.

FRANCESCO SIMONE (Member, IEEE) received the B.Sc. degree in mechanical engineering and the M.Sc. degree in mechanical engineering with a specialization in industrial production. He is currently pursuing the Ph.D. degree in industrial and management engineering with the Sapienza University of Rome, Italy. He is also a former Contract Researcher at the Department of Mechanical and Aerospace Engineering, Sapienza University of Rome. His current research interests include systems theory applications on complex systems' resilience and reliability through the use of simulation tools and AI techniques.

ANTONIO JAVIER NAKHAL AKEL received the double B.Sc. degree in mechanical engineering, and the M.Sc. degree in mechanical engineering with specialization in industrial production, from the Sapienza University of Rome, Italy, where he is currently pursuing the Ph.D. degree in industrial and management engineering. His B.Sc. studies were sponsored by the Central University of Venezuela, Venezuela and the Sapienza University of Rome. He is also a former Contract Researcher at the Department of Mechanical and Aerospace Engineering, Sapienza University of Rome. His current research interests include resilience management for industrial socio-technical systems in critical complex contexts through the use of business intelligence tools and artificial intelligence techniques.

GIULIO DI GRAVIO is a Full Professor of industrial engineering and operations management at the Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Italy. He has published about 120 papers in academic journals and conference proceedings. His research interests include the analysis and design of industrial, organizational, and enterprise network systems, collaboration and coordination of supply chains, performance and risk management, and resilience management.

RICCARDO PATRIARCA received the B.Sc. degree in aerospace engineering, the M.Sc. degree in aeronautical engineering, and the Ph.D. degree in industrial and management engineering. He is currently a Tenure Track Assistant Professor at the Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Italy. He has published widely (about 100 papers published in academic journals and conference proceedings) on methodological and epistemological aspects of risk, safety, resilience management, and operations management at large. He aims to make systems safer and resilient when, and especially before, things go awry. He received the Doctor Europaeus Certification during his Ph.D. degree.

Open Access funding provided by 'Università degli Studi di Roma "La Sapienza"' within the CRUI CARE Agreement