Cybersafety: A System-Theoretic Approach to Identify Cyber-Vulnerabilities & Mitigation Requirements in Industrial Control Systems

Shaharyar Khan and Stuart Madnick, *Member, IEEE*

Abstract—Recent cyber-physical attacks, such as *Stuxnet, Triton* etc., have invoked an ominous realization about the vulnerability of critical infrastructure, including water, power and gas distribution systems. Traditional IT security-biased protection methods that focus on improving *cyber hygiene* are largely impotent in the face of targeted attacks by advanced cyber-adversaries. Thus, there is an urgent need to analyze the safety and security of critical infrastructure in a holistic fashion, leveraging the *physics* of the cyber-physical system. System-Theoretic Accident Model & Processes (STAMP) offers a powerful framework to analyze complex systems; hitherto, STAMP has been used extensively to perform safety analyses but an integrated safety and cybersecurity analysis of industrial control systems (ICS) has not been published. This paper uses the electrical generation and distribution system of an archetypal industrial facility to demonstrate the application of a STAMP-based method – called *Cybersafety* – to identify and mitigate cyber-vulnerabilities in ICS. The key contribution of this work is to differentiate the additional steps required to perform a holistic cybersecurity analysis for an ICS of significant size and complexity and to present the analysis in a structured format that can be emulated for larger systems with many interdependent subsystems.

Index Terms—CPS security design, industrial control system, STAMP, system security, cyber-physical damage

1 Introduction

Whereas cyber-physical attacks targeting automobiles, medical devices and other systems embedded with computers have the potential to cause considerable damage to individuals or small groups of people, a cyberattack targeting critical infrastructure ICS can impact a large number of people over a vast geographical area. This is why such attacks are considered a matter of *national security* [1].

The 2009 Stuxnet cyberattack that partially destroyed a third of the centrifuges at a uranium enrichment facility in Natanz, Iran, ushered a new era in cyber warfare [1]. Since then, several attacks around the world including the Ukraine power grid attacks (in 2015 and 2016), *Triton* attack targeting *safety-instrumented systems* at a Saudi industrial facility in 2017 etc., have demonstrated not only the unprecedented *capabilities* of such attacks on causing widespread disruption and/or destruction [1], [2], but the *willingness* of nation-states to exploit such vulnerabilities in an opponent's critical infrastructure.

Therefore, there is an urgent need to reevaluate the *safety* of critical infrastructure industrial control systems in the context of *cybersecurity* threats to such systems. The traditional approach to protecting such systems is to undertake a risk-based, technical perspective that is biased by information security concerns. Such IT security-biased protection methods that narrowly focus on improving cyber hygiene are only successful against indiscriminate, non-targeted attacks – but remain largely impotent against targeted attacks by advanced cyber adversaries [3].

In reality, *security*, like *safety*, is an emergent property of the system where the interactions of simple components produce complex behaviors – underscoring the need for a *systems perspective* of the security problem.

The unique contribution of this paper is to present the results of a cybersecurity analysis of an archetypal ICS using a system-theoretic method based on the STAMP framework [18]. Hitherto, STAMP has been used extensively across many industries to perform safety analyses but an integrated safety and cybersecurity analysis of ICS has not been published. In this paper, we analyze the electric generation and distribution system of a small-scale industrial facility. The paper aims to provide a repeatable method which can be emulated to analyze larger industrial control systems.

Specifically, the analysis highlights how an attack on the digital automatic voltage regulator (AVR) of a generator could destroy the generator in the matter of a few seconds as a result of hazardous control actions and how redesigning the control structure through fail-safe design, changes in processes and procedures and social controls (such as policy, culture, insurance incentives etc.) could prevent such a loss.

For instance, among other things, it is shown how in the event of an attack on the AVR, the inclusion of a relatively inexpensive relay (~$6,000 [4]) could avert the loss of a

- *Shaharyar Khan is with the Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. E-mail: shkhan@mit.edu.*
- *Stuart Madnick is with the Sloan School of Management and School of Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. E-mail: smadnick@mit.edu.*

Manuscript received 9 January 2020; revised 14 April 2021; accepted 24 June 2021. Date of publication 29 June 2021; date of current version 2 September 2022.
(Corresponding author: Shaharyar Khan.)
Digital Object Identifier no. 10.1109/TDSC.2021.3093214

turbo-generator (~$11M [5]) and subsequent outage costing several million dollars in repairs and lost revenue. It also provides several realistic scenarios that illustrate how the interdependencies of the controlled process could be exploited to enable such an attack. Section 2 provides a literature review about the application of systems theory to cybersecurity. Section 3 provides a brief overview about the *Cybersafety* method. Section 4 describes the key features of an archetypal industrial plant that the method was applied to while Section 5 describes the bulk of the analysis. A discussion of the results, along with some proposed mitigation requirements is provided in Section 6 followed by a short conclusion in Section 7.

2 Literature Review

Traditional approaches to protect cyber-physical systems are often strongly biased by practices and design principles prevalent in the information security world. These protection mechanisms and principles broadly include authentication, access control, firewalls, intrusion detection, antimalware, application whitelisting, flow whitelisting, cryptography, integrity verification, survivability etc. *Loukas* [1] and *Cardenas et al.* [6] support the view that traditional protection mechanisms in cyberspace are largely applicable to cyber-physical systems. However, they note that important differences exist in implementation and effectiveness; some of these are described next.

First, for cyber-physical systems, *availability* and *integrity* of information is more crucial than confidentiality of information. Second, for intrusion detection in cyber-physical systems, sensor data from the physical space is an important input, unavailable to IT systems which rely purely on cyberspace metrics. Third, an understanding of the consequences of an attack in the physical world is required to design a protection scheme for defense-in-depth of the cyber-physical asset. And fourth, conventional security policies (such as *patching*) may in fact increase potential vulnerabilities, rather than decreasing them [7] – to elaborate this point, note that for air-gapped *Supervisory Control and Data Acquisition* (SCADA) systems, the *'air-gap'* improves cybersecurity as it limits opportunities for remote attacks. Paradoxically, however, this also means that the devices cannot be automatically updated with malware signatures (blacklists) and operators must manually install any updates on each isolated device, thereby increasing the risk of *cross-contamination* due to frequent manual updates [7].

2.1 Safety Focused Approaches

Since one of the primary concerns with security of cyber-physical systems is its impact on system safety, a number of hazard analysis frameworks and methods traditionally used for safety analyses have been adapted for security analyses.

For instance, *Schmittner et al.* [8] extended Failure Modes and Effects Analysis (FMEA) [9] to include security consideration, by including vulnerabilities, threat agents and threat modes as inputs for determining failure causes. The extended method is called Failure Mode, Vulnerabilities and Effects Analysis (FMVEA). Likewise, *Steiner and Liggesmeyer* [10] proposed an extension of Fault-Tree Analysis (FTA) [11] by modeling attacker's intentions in the analysis; the method is known as the Extended Tree Analysis (ELT) [12].

Despite these advances, there are inherent limitations of these methods. For instance, while FMEA is well-suited for evaluating individual component failures and providing reliability information, it is limited in its use as a safety tool because it considers single item failures without considering failures due to component interactions [13], [14].

Likewise, *Xu et al.* [15] argue that FTA is limited in its analysis of human factors, organizational and extra-organizational factors. It also fairs poorly as the complexity of the system increases [15]. *Leveson* [16] argues against the use of probabilistic risk analyses (i.e. the underlying framework for FTA) over system design analyses to improve system safety due to the inherent difficulty and uncertainty in assigning probabilities to design and manufacturing flaws.

According to *Dunjó et al.* [17], the systems-based Hazard and Operability (HAZOP) Analysis [13], [14] lies in between FTA and FMEA. *Friedberg* [18] argues that over the years, researchers have tried to formalize HAZOP to achieve objective and quantifiable results, *"but all approaches to quantify results have led back to the use of FTA"*.

2.2 System-Theoretic Accident Model and Processes (STAMP)

An alternative to performing joint analysis of safety and security using extended versions of traditional hazard analysis methods (such as FTA/FMEA etc.), is to use the perspective of modeling using *systems theory*. Leveson [19], [20] developed a framework to understand causes of accidents using systems theory. This framework is called STAMP (System-Theoretic Accident Model and Processes).

STAMP treats accidents as a *'dynamic control problem'* emerging from violation of safety constraints rather than a *'reliability problem'* aimed at preventing component failures. Several analytical methods have been developed based on the STAMP framework such as STPA, CAST etc.

STPA is an acronym for System-Theoretic Process Analysis; it is a forward-looking approach for identifying hazards in complex systems [19], [20]. Similar to STPA, but looking backwards, CAST (Causal Analysis using Systems Theory), is used to identify causal factors for past events or accidents using the STAMP framework [19], [20].

In his thesis, *Thomas* [21] provides a mathematical model underlying STPA and a method to perform the analysis systematically which enables a more rigorous analysis with more objective results. Since its creation, the STPA method has been applied to a wide variety of industries and *safety* use-cases. It has been used in the automotive industry [22], automation and workplace safety [23], nuclear power plants [24], ship navigation [25], medical applications [26] etc.

Laracy [27], [28] recognized the similarities between safety and security and proposed an extension of STAMP to security problems of critical infrastructure, such as the Air Transportation System. This approach was called STAMP-Sec [27].

Salim [29] performed the first documented cybersecurity analysis using the STAMP-based CAST method by analyzing the TJX Cyberattack; this was the largest cyberattack in history (by number of credit cards) when announced in 2007 and cost TJX $170 million. *Nourian and Madnick* [30]

furthered this research by applying the CAST method to analyze the infamous Stuxnet cyber-physical attack. The notion of combining safety and security analysis into an integrated approach for hazard analysis was presented in a concept paper by *Young and Leveson* [31] and the method was called STPA-Sec.

Schmittner et al. [32] highlight some of the limitations of applying STPA-Sec and propose extensions of the STPA-Sec methodology. This includes alignment of terminologies between safety and security and provision of guidewords to elicit scenarios due to malicious actions in the final step of the analysis.

An extension to STPA-Sec, *Friedberg et al.* [18] present an analytical methodology that combines safety and security analysis, known as STPA-SafeSec. The core contribution in this work [18] is the mapping of the abstract control layer used in the STPA analysis to physical components for which security constraints are defined. Note that the essence of the STAMP framework is in the functional control structure, process models and constraints; by introducing a physical layer, the analysis becomes much more complicated with an inadvertent focus on component-level vulnerabilities. Even for relatively simple systems (such as the synchronous islanded generation use case described by *Friedberg et al.* [18]), the analysis becomes laborious. In addition, STPA-SafeSec introduces general integrity/availability threats (command injection, command drop etc.) as a guidance for mapping causal factors between the control and component layer. However, this reduces the scope of the analysis to technical components only. As opposed to STPA-SafeSec, Cybersafety attempts to capture the STAMP ideology focusing on vulnerabilities emerging from violation of constraints for components as well as due to component interactions (both direct and indirect) at the functional level throughout the larger socio-technical system.

The Idaho National Lab (INL) [3] developed a novel approach called *Consequence-driven Cyber-informed Engineering* (CCE) that is also inspired in part by work done by *Leveson* [19], [20]. Similar to STAMP, CCE is a top-down approach that is consequence driven and considers system interdependencies. However, whereas STAMP is focused on holism and dynamic control, CCE resorts to analytic reduction early on in the analysis (by undertaking a system-of-systems breakdown). This work is still in its early development phases and information about the method and its implementation is scarce.

While STPA (and STPA-Sec, etc.) have been proposed as tools to identify and help mitigate cyberattacks on industrial control systems, this analysis is the first (and perhaps, only) extensive example demonstrating that it can actually work [33]. Our literature search did not reveal any detailed published work documenting the application of STPA-Sec to industrial control systems or power generation plants with the exception of fictionalized educational examples. In this paper, we incrementally refine the STPA-Sec method into a robust, systematic and repeatable set of steps by demonstrating its application to the electric generation and distribution system of a small-scale industrial facility. We refer to this *focused* approach for identification and mitigation of cyber-related vulnerabilities in ICS as *Cybersafety* which is described in the following section.

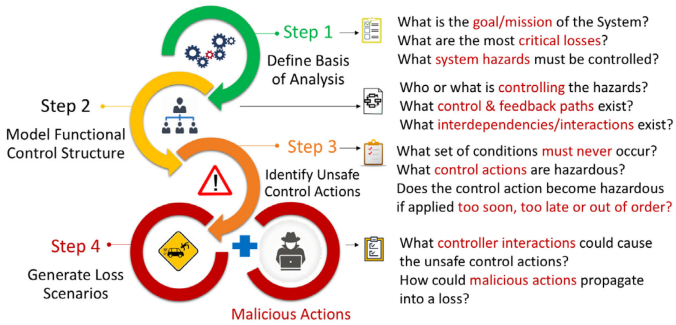

Fig. 1. Overview of the cybersafety method.

The key contributions of this work include elaboration of steps required to analyze an industrial control system of significant complexity and size, with diverse functionality, in the context of cybersecurity. It also includes specifying the logical thought process to identify system-level hazards and enumeration of steps to repeatably develop the functional control structure at a level of abstraction that is sufficient to enable a comprehensive analysis. In addition, it outlines the method to identify process model variables for controllers considering system interdependencies and a formal approach for generating loss scenarios and rich causal factors that transcend direct causality and attempt to identify flaws in the system emerging from controller interactions.

3 Cybersafety – A Systems Perspective of Security

The basic steps in the *Cybersafety* method are identical to STPA and summarized in Fig. 1. A brief description of the main steps and the key improvements is summarized next.

Step 1: Define the basis of the analysis by identifying worst-possible outcomes for the system as well as those system states (i.e. system hazards) that if not controlled would result in the worst-possible outcomes. In the *cybersafety* method, we have added a step to identify *critical functions* that enable the target system to achieve its *goal* or *mission*. This enables deriving the system hazards by focusing on the critical functions of the system which is more meaningful for developing the hierarchical functional control structure in Step 2. We have also added a step to explicitly identify *interdependencies* of the target system.

Step 2: Develop a hierarchical functional control structure to model the controllers and their interactions that together are intended to enforce safety and security constraints on the system. In the *cybersafety* method, we have outlined steps that ensure the completeness of the functional control structure based on system-hazards identified in Step 1. In addition, we extend the functional control structure beyond the target system to include interactions with the *environment* – based on system interdependencies identified in Step 1.

Step 3: Identify control actions that could be hazardous and lead to system disruption or damage. In the *cybersafety* method, we additionally define logical steps to identify variables for the process model; this implicitly accounts for system interdependencies identified earlier.

Step 4: Generate loss scenarios leading to the unacceptable worst-possible outcomes identified in Step 1. In a

departure from traditional STPA analysis, malicious actions are also considered as causal factors leading to system hazards. Two categories of causal scenarios are considered which include [19]:

a. Scenarios where an unsafe control action is issued
b. Scenarios where a safe control action is provided but not followed or executed properly

In the *cybersafety* method, we also provide an approach for identifying rich causal factors that go beyond direct causality by focusing on flaws including, process/mental model flaws, structural flaws, contextual factors, coordination and communication flaws, and system dynamics factors that enable the loss scenario. As a final step, new functional requirements and mitigation strategies are defined that would prevent the worst-possible outcomes identified in Step 1.

With *Cybersafety* (similar to STPA-Sec [31]), instead of focusing on *threats* from adversaries which are beyond the control of the system, security efforts are focused on controlling system vulnerabilities *internal* to the system, which the defender has control over. This enables preventing disruptions from not only known threats, but also unknown threats, such as insider-attacks [31]. In contrast to traditional security approaches where vulnerabilities are a function of known threats, in *cybersafety*, vulnerabilities are a function of system design. The concept is to engineer out a solution in the design of the control structure of the system, so that the system becomes inherently more safe and secure. Here, the system is viewed as a collection of dynamically interacting hierarchy of controllers; making the success of an attack contingent on the ability of the controllers to detect an anomaly and restore the controlled process to operate within certain defined constraints.

The *Cybersafety* method, being based on the STAMP framework, is inherently a *qualitative method* that is focused on identifying not what is *likely* to go wrong, but what can *possibly* go wrong. *Leveson* [16] (referring to safety) argues against the use of quantitative approaches based on *severity* and *likelihood* alone because of lack of availability of good estimates for *likelihood*. Complex systems (especially those that do not have historical precedence) have a high degree of uncertainty associated with them which makes the process of estimating likelihood error prone. *Leveson* [16] states that the addition of a quantification step *"introduces so many uncertainties and inaccuracies that it undermines any safety-related decision process based on it"*.

Several industrial accidents lead credence to this argument. For instance, the Lithium-ion batteries on the Boeing 787 caught fire twice within 50,000 flight hours of operation, when they were certified on the basis of an estimate that there would be no more than 1 fire in 10,000,000 flight hours [16]. In contrast to a risk-based approach, a design analysis-based qualitative approach focuses on enumerating hazards and mitigating them in the design of the system.

The same argument can be extended to security analyses as well. Given that industrial environments have become increasingly complex and coupled, with a juxtaposition of old and new software-based control technology (increasing the uncertainties in the system), it is crucial to advance use of qualitative analyses as a means for making safety and security-related decisions.

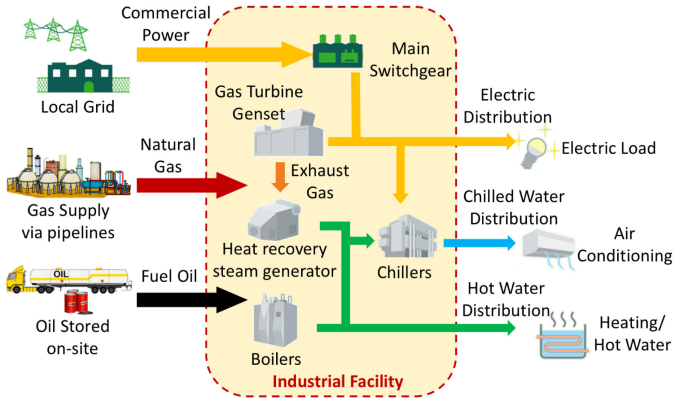

Fig. 2. The plant - A microcosm energy facility.

Note that there is no such thing as a *'secure'* system – only *'securer'* systems. The goal of the *Cybersafety* method is to undertake a top-down, systems perspective of the security problem to shed light on blind spots that may not be visible via other methods due to their narrow, component-centric risk-based approach. This method is not intended or claimed to make the system 100 percent secure (doubt any approach could provide such a measure of confidence); rather, it is intended to enhance a system's security posture than it otherwise would be. As a potential extention in future work, we would consider complementing the qualitative analysis proposed here with a quantitative evaluation to improve the robustness of the method.

4 The Plant

The industrial facility that is the subject of this study is an archetypal energy facility with upstream operations that include delivery of fuel (both natural gas and fuel oil) to the plant along with a tie-line connection to the local utility grid as well as downstream operations that include distribution of electricity, steam and chilled water to the facility. The plant operates a 21 MW ABB (GT10) gas turbo-generator that provides electricity to the facility; waste heat from the turbine is directed to a Heat Recovery Steam Generator (HRSG) to produce steam. The steam along with other gas/oil-fired water-tube boilers is used for heating and other functions such as driving steam-driven chillers. The chilled water supply from steam-driven chillers is complemented by several electric-driven chillers to meet demand. A schematic of the plant's equipment and operations is shown in Fig. 2. The key processes and equipment that make up the plant's generation and distribution system are summarized next. Detailed descriptions of each equipment and process is provided by *Khan* [34].

The power generated by the gas turbine meets only about 60 percent of the facility's electricity demand; the shortfall is drawn from the local utility tie-line. The industrial plant is served from the local utility by six 13.8 kV service connections feeding into the main switchgear in parallel with the gas turbine which also produces power at 13.8 kV. The switchgear consists of various switching and protection devices including switches, circuit-breakers, reclosers and fuses [34]. The Medium Voltage (MV) circuit breakers operate when directed remotely by the operator or digital protective relays to *open* or *close*. Many different types of digital

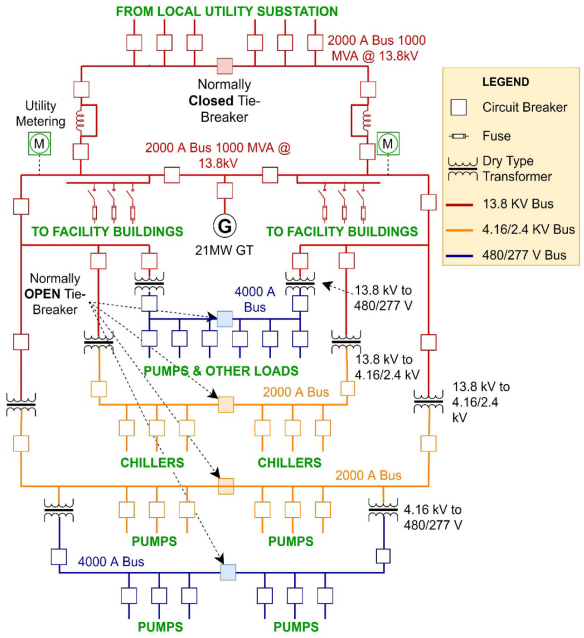

Fig. 3. One-line diagram of plant's electrical distribution system.

protective relays (overvoltage, over-current, directional etc.) protect different parts of the distribution network by opening/closing the required circuit breaker(s) to isolate equipment, feeders, buses etc., using feedback from sensors (such as current transformers (CT) or potential transformers (PT)) and pre-set control algorithms.

The primary electric distribution system at the plant is configured as a loop system designed with redundancy throughout the facility to provide a high-level of service continuity as shown in Fig. 3.

A Distributed Control System (DCS) is implemented to control and manage the devices that affect the plant's electric generation and distribution, along with other plant equipment, including, boilers, chillers and ancillary equipment. The DCS is integrated with the *Turbine Controller* that manages on-site electricity generation, adjusting its output to meet the industrial facility's active and reactive power demand as directed by the operator. The operator, in turn, uses an *Energy Management System* (EMS) from an external vendor as guidance to most optimally assign setpoints for plant equipment, considering electricity and gas price fluctuations.

The electric generation and distribution system is a complex system with many components interacting in indirect ways. We will now demonstrate the application of the *Cybersafety* method to logically and systematically identify vulnerabilities in the system emerging as a result of interactions between the various components of the target electric generation and distribution system.

5 Analysis

5.1. Define Basis of the Analysis – Step 1

Cybersafety is a top-down, consequence-driven approach that begins by establishing the boundaries of the system by defining the goal of the system and identifying the critical functions required to achieve that goal along with unacceptable system-level losses. The *system problem statement* provides a convenient framework for establishing the goal and

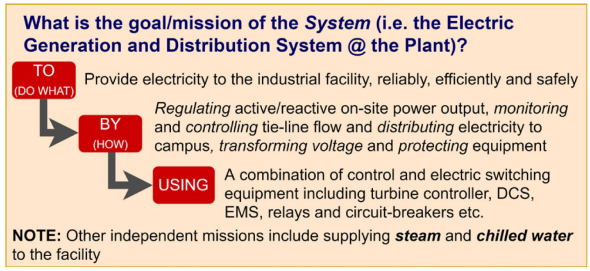

Fig. 4. System problem statement.

critical functions of the system as shown in Fig. 4. By defining the *critical functions* in the *system problem statement*, we can focus on those losses and hazards that are most critical to the success of the mission or goal of the target system.

1. *Unacceptable System-Level Losses.* An unacceptable system-level loss is any condition that is unacceptable from the primary stakeholder or mission owner's perspective. The unacceptable system-level losses for the electric generation and distribution system are itemized in Table 1. The list is deliberately kept high-level and has been defined in terms of the system rather than individual component losses. This is done to manage complexity – by starting with a short list at a high-level of abstraction, one can be more confident

TABLE 1
Unacceptable Losses, Hazards and Constraints

Losses	Hazards	Constraints
L-1: Death, dismemberment or injury to personnel	H-1: System is operated beyond normal operational limits [L-1, L-2, L-3, L-4] H-1.1: Mechanical parameters (speed, ramp rates, temp., pressure, vibration as applicable) exceed normal operational limits H-1.2: Electrical parameters (current, voltage, frequency) exceed normal operational limits	SC-1: System must prevent operation beyond normal operational limits
L-2: Physical damage to critical equipment	H-2: System does not adequately control active & reactive power [L-2, L-3, L-4] H-2.1: Violation of power quality metrics H-2.2: Generation does not meet system load H-2.3: Unable to achieve synchronization	SC-2: System must adequately control active and reactive power to meet power quality metrics, system load and maintain synchronization
L-3: Loss of mission i.e. inability to deliver electricity to campus	H-3: System does not correctly sequence operations [L-1, L-2, L-3, L-4] H-3.1: Out-of-sync re-closure H-3.2: Operation without permissive function H-3.3: Out-of-order switching operation	SC-3: System must correctly sequence operations including re-closures & switching operations and ensure permissive functions are not violated
L-4: Economic loss due to an electrical event (including capital cost or operational cost)	H-4: System does not correctly isolate faults [L-1, L-2, L-3, L-4] H-4.1: Faulted area is not isolated H-4.2: No-fault are is isolated	SC-4: System must correctly and promptly isolate faults
	H-5: System does not adequately transform voltage [L-1, L-2, L-3, L-4] H-5.1: Operation in overloaded condition H-5.2: Operation in overexcited condition	SC-5: System must adequately transform voltage SC-5.1: System must shed load to prevent overloading SC-5.2: System must prevent operation in overexcited state

about completeness of the analysis because each of the longer lists of causes can be traced back to one or more of the small starting lists (and vice versa) [19].

2. *System-Level Hazards and Constraints*. In a complex system, due to the complex nature of interactions between the components of the system, it is not always feasible or possible to predict the exact nature of interactions of each component of the system at every moment in time. The system as a whole, however, exhibits emergent behaviors which must be constrained to operate within certain defined limits. Certain conditions or system states move the system beyond these safe limits (resulting in losses); these conditions or system states are called *system-level hazards* [19].

Leveson [19] argues that the hazards must be defined in terms of the overall system behavior – not components. However, for a complex system, with multiple independent functions it is difficult to directly define hazards in terms of the system which communicate any meaningful information. All efforts to define hazards in terms of the system (such as a power generation plant) were found to be in vain; either the hazards were too high-level to provide any meaningful information about the system or they ended up being defined in terms of components.

Instead, it was discovered that if the *critical functions* identified in the *system-problem-statement* are inverted, a coarse list of system-level hazards can be defined in terms of system functions. Focusing on each of the high-level hazards in the coarse list, a more refined list of hazards (or unsafe system states) can be generated (as listed in Table 1) which can inform the development of the *functional control structure* in Step 2 of the method. Note that since the key *functions* that enable the system to achieve its primary goal/mission are utilized to generate the list of hazards, we can be more confident about the coverage and accuracy of the identified hazards. In addition, we validated the list of hazards by plant engineers at the target facility.

For each hazard, constraints must be defined which prevent the hazard from translating into system-level losses. As a first approximation, inverting the list of hazards, yields a list of constraints as shown in Table 1. Progressing through the analysis, this list of constraints is systematically refined, ultimately, resulting in a set of functional requirements to protect against specific loss scenarios in Step 4.

5.2 Model the Functional Control Structure – Step 2

The previous subsection concluded with a definition of constraints that prevent the system hazards from propagating into unacceptable losses. In turn, the system hazards are derived from critical functions that enable the system to achieve its goal. In this subsection, we model how these constraints are enforced on the system via a hierarchy of controllers known as the functional control structure. At its most fundamental level, the functional control structure models control loops consisting of controlled processes and controllers. Fig. 5 shows the high-level functional control structure for the electric generation and distribution system.

Recognizing the processes that must be controlled to prevent the system hazards, the high-level function control structure is carefully refined to add more detail. Fig. 6 shows a refined version of the functional control structure.

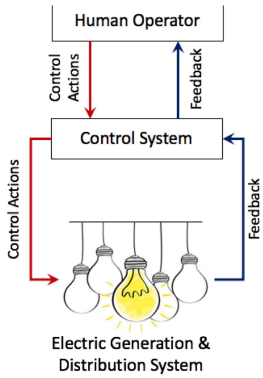

Fig. 5. High-level functional control structure.

The figure shows a system of interacting controllers, primarily enforcing constraints on two controlled processes – *on-site generation* and *on-site distribution* function (i.e. electricity distribution through circuit-breaker control). The controllers for on-site generation include the turbine controller, the automatic voltage regulator as well as the synchronization unit (regulating relay). The controllers for the switching function include protective relays as well as automatic-load transfer switches.

Supervisory controllers include the DCS as well as the Real-time Automation Controller (RTAC). RTAC provides automatic load-shedding along with automated system stability functionality. The supervisory controllers are managed by operators, who in turn are controlled via work instructions by plant engineers as well as through policies enforced by the plant management. By recursively asking the question who is controlling what, the higher-level controllers, beyond the human operator can be identified. This hierarchical modeling provides insights about the flow of control in the system which can be leveraged, later in the analysis, to derive more effective mitigation strategies.

Thus far, the focus has been on understanding the control structure for the electric generation and distribution system. However, the electric generation and distribution system cannot be studied in isolation since it has a strong interdependency with other systems both inside the plant as well as outside the plant (with systems it has control over as well as systems it does not have control over). *Rinaldi et al.* [35], describe a robust approach for identifying system interdependencies by considering *Physical, Cyber, Geographical* and *Logical* interdependencies systematically. Following this approach, the dependencies and interdependencies of the electric generation and distribution system of the plant are identified in Table 2. The table illustrates the dependency of the electric distribution system on natural gas, fuel oil, water as well as local electric utility distribution systems.

5.3 Identify Unsafe Control Actions – Step 3

The next step in the Cybersafety method is to identify *Unsafe Control Actions*. Note that a particular control action in of itself is not unsafe, rather the context in which it is performed, makes it *safe* or *unsafe*. We begin by identifying the primary functions, safety responsibilities and associated control actions for each controller in the functional control structure as presented in Table 3.

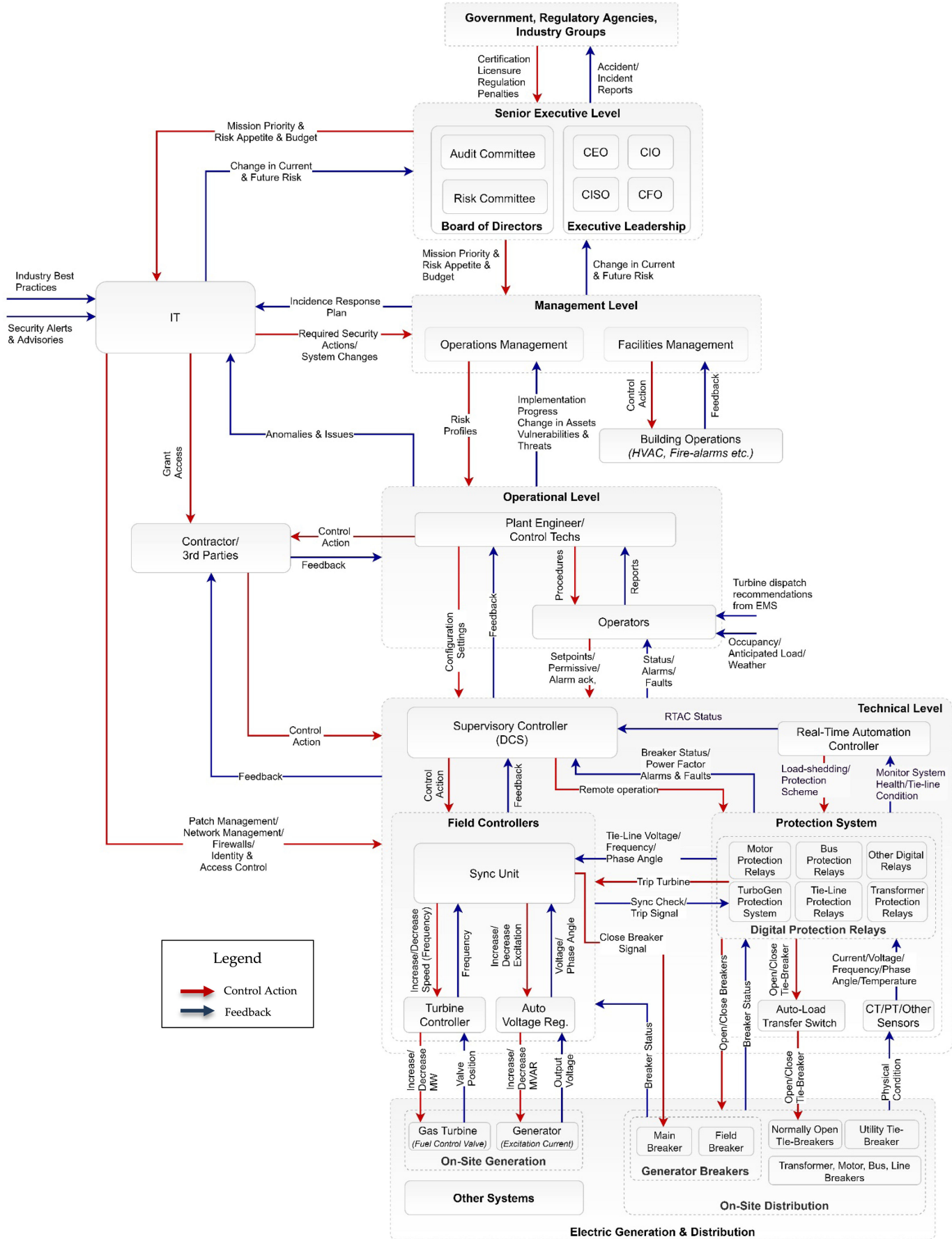

Fig. 6. Detailed hierarchical functional control structure.

TABLE 2
Interdependencies of the Target System

Physical (Inputs/Outputs) *An infrastructure is physically dependent if there is a functional & structural linkage between the input & output of two assets: a commodity produced by one infrastructure is required by another infrastructure for its operation* • Natural gas at a min. pressure of 300 psig • Fuel oil for backup stored on site • Lube oil for turbine, generator • Water for cleaning of compressor, emissions control • Purified air • Electricity from the local grid
Cyber (Informational) *An asset has a cyber dependency if its operation depends on information transmitted via electronic or informational links* • Communication with external contractors/vendors for system monitoring via data link • Energy Management System (EMS) for gas turbine, chillers, boilers throttle settings • Plant Information (PI) system for business operations and engineering troubleshooting
Geographical (Logical) *Assets are geographically dependent if an event in the local environment can create changes in those assets' state of operations. A geographic dependency occurs when elements of infrastructure assets are in close spatial proximity* • Proximity of gas turbine to boilers and chillers • On-site fuel storage
Logical *An infrastructure is logically dependent if its state of operations depends on the state of another infrastructure via a mechanism that is not a physical, cyber or geographic connection. Logical dependency is attributable to human decisions and actions* • No logical interdependencies can be identified at this time

Next, the *process model* for each controller is determined; this is the model that the controller uses to determine what control actions are *safe* or *needed* in order to keep the controlled process within certain limits. Importantly, the process model provides context to the controller's decision-making process by focusing on environmental factors that can influence the state of the controlled process. The process model is a potential target for the attacker as it can be leveraged by the *attacker* to cause hazardous control actions to be issued by the controller.

The variables that must be considered in formulating the *process model* can be identified by evaluating the following: 1) the *state* of the process that is being controlled, 2) the definition of *system hazards* related to the controlled process, and 3) the *environmental conditions* that would cause the controller to change its state or the interdependent processes that would be affected as a result of a change in state. The *process model* for one controller, the *Digital Automatic Voltage Regulator* (AVR), is presented in Table 4.

Once the process model variables have been identified, unsafe control actions (UCA) can be recognized by enumerating each potential combination of relevant process model variables and examining whether the issuance of the control action in that system state would be hazardous [36]. Several UCAs for the AVR are listed in Table 5 along with a partial list of UCAs for the Plant Operator and the Plant Management. The important thing to note here is that each UCA is defined in terms of the context of a system state i.e. under certain conditions, *nominally* safe control actions become hazardous. Also note that each UCA is tied back to a system-level hazard identified in Step 1.

For each UCA, a corresponding safety/security constraint must also be defined. For example, a potential safety constraint for UCA-AVR-5 could be defined as follows: *"AVR must not increase excitation when generator frequency decreases below synchronous speed (to prevent high V/Hz (overfluxing) during islanded or grid operation). [SC-1]"*

Note that just how UCAs are refinements of system-level hazards defined in Step 1, these safety/security constraints are also refinements of system-level constraints defined in Step 1 of the analysis. These constraints on their own provide an excellent source of first-order requirements to protect the system. However, it is important to recognize that these are component level constraints; further refinement of these constraints, considering the interactions between the components is provided in the next subsection.

From Table 5 it is evident that the AVR performs a critical function in the stabilization of voltage and maintenance of power quality metrics. Improper operation of the AVR can significantly damage the generator in a matter of a few seconds resulting in *"expensive repairs, several months of forced outage and loss of production"* [37], [38], [39]. The UCAs summarized in Table 5 are discussed next.

5.3.1. Overexcitation

Overexcitation occurs whenever the ratio of the voltage to frequency (V/Hz) applied to the terminals of the generator exceed design limits [40], causing high flux density levels (i.e. overfluxing) in the generator core. *Mozina* [41] states, "at high flux levels, the magnetic iron paths designed to carry the normal flux saturate, and flux begins to flow in leakage paths not designed to carry it". The resulting fields can cause overheating of the stator core iron, and under severe overexcitation conditions, result in the partial or complete destruction of the stator core's insulation [41]. Typically, generators are designed to handle a full load field with no load on the machine for 12 seconds before the stator iron laminations become overheated and damaged [42]. The overexcitation conditions can be caused by overvoltage, under-frequency or a combination of both [41] (UCA-AVR-5, -7, -8).

This condition can also be a result of an operating error during manual regulator control or sudden load rejection. Additionally, if the unit is connected to a capacitive load and there is a sudden loss of load, leading reactive current would flow into the machine. If this reactive current flow is close to the minimum excitation limit of the AVR, the regulator will boost the excitation in an attempt to reduce the reactive current flow into the machine, increasing the terminal voltage of the machine, possibly causing overexcitation [43] (UCA-AVR-11).

5.3.2 Excessive Field Current – Field Overexcitation

Another related condition is field overexcitation; this condition occurs when the rotor field current is raised beyond its normal limits. Such a condition can result in excessive heating of the rotor windings due to field overcurrent. This

TABLE 3
Partial List of Controllers, Safety Responsibilities & Control Actions

Controller	Function Performed	Safety & Security Responsibilities	Control Actions
Digital Automatic Voltage Regulator	Maintain generator terminal voltage at a set value under varying load and operating conditions; absorb or deliver reactive power for PF control	• Control generator terminal voltage through adjustment of field current • Maintain system voltage when operating in islanded mode	• Control excitation (increase/decrease)
Turbine Controller	Maintain desired fuel flow to the turbine to achieve operator provided setpoints	• Control turbine active power • Protect against overspeed, overtemperature, excessive vibration, loss of flame, loss of lube oil and other hazardous conditions • Prevent equipment damage by correct sequencing of operations	• Control fuel valve (increase/decrease) • Start/Stop pumps, valves for safe operation (lube oil, aux lube oil etc.)
Digital Protection Relays	Microprocessor-based relay that analyzes power system voltages, currents or other process quantities for the purpose of detection of faults	• Remove any individual component of a power system when it suffers a fault that might result in damage to property or unsafe conditions	• Trip breaker
Distributed Control System (DCS)	Provide operator with supervisory control and monitoring of all automated controllers distributed throughout the plant	• Raise alarms and faults for system abnormalities • Aggregate data and provide accurate information • Ensure necessary auxiliary equipment (such as pumps/valves) are open/closed	• Start/stop auxiliary equipment
Operator	Perform day-to-day tasks to run equipment including the gas turbine, boilers and chillers in response to real-time demand variations from the facility	• Monitor system operation for abnormalities • Emergency shutdown of equipment • Respond to alarms/faults and take corrective actions • Provide permissive functions/command overrides • Report Anomalies	• Shutdown equipment safely during emergencies • Manual override commands to keep equipment within safe limits
Plant Engineer	Act as the technical lead for plant operations; write operating procedures, ensure compliance with procedures; respond to incidents	• Certify design, equipment & procedures for safe operation; develop operating procedures • Ensure procedural compliance & training • Respond to incidents • Report issues, anomalies to management	• Equipment & design certification • Approve operating procedures • Provide technical specifications and requirements to contractors/ vendors • Change control logic/configuration
Information Technology (IT) Security	Protect the organization's confidentiality and integrity of processed data, provide availability of systems and data, and assurance that the system will meet its design goals	• Ensure the plant's IT systems are free from malicious code, vulnerable systems are identified, reported and adequately mitigated	• Vulnerability Scanning • Incident response assistance • Access Control • Patch Management
Plant Management	Ultimately responsible for safe operation of the plant	• Ensure plant complies with all safety/security regulations • Ensure engineers and operators have adequate resources and training for safe operation of the plant • Ensure budget allocations are consistent with risk profile	• Allocate budget • Set priorities

condition is different from the overfluxing condition described earlier since one is caused by a high V/Hz ratio while the other is caused by an overcurrent condition (UCA-AVR-10).

5.3.3 Overvoltage

Overvoltage occurs when the levels of electric field stress exceed the insulation capability of the generator stator windings [40]. This condition is again distinct from overfluxing since a high voltage with a proportionally high frequency would not cause an overfluxing event, but it would result in an overvoltage condition (UCA-AVR-6). Excessive voltages can damage and break down stator insulation in the machine leading to a fault [43]. It can also stress insulation in other connected components such as transformers, bushings and surge arrestors.

TABLE 4
System Variables for the AVR and their possible values

#	AVR Process Model Variables	Process Model States
1	Excitation Level (Gen. Terminal Voltage)	Below | At Setpoint | Above
2	Generator Breaker Status (Islanded vs. Grid)	Open | Closed
3	Gen. Operating Point (Capability Curve)	Within Limits | Outside Limits
4	Grid Voltage	Within Limits | Outside Limits
5	Frequency	Within Limits | Outside Limits
6	Reactive Power Demand	Leading | Lagging
7	Turbine Trip Status	Tripped | Not Tripped

5.3.4 Under-excitation/Loss of Field

In contrast to overvoltage, not providing enough excitation, can also be hazardous. When not synchronized to the gird (for instance, during startup) if the AVR does not increase excitation to match generator's terminal voltage with the system voltage (grid) it cannot be synchronized to the grid (UCA-AVR-2). On the other hand, when connected to the grid, excitation controls the reactive power fed into the power system, which in turn dictates the plant's power factor. When the field excitation is less than what is required to maintain the generator's terminal voltage at or above the grid voltage, reactive current flows into the generator stator windings, which can cause overheating of the stator core and insulation damage; this condition is called under-excited power factor operation. Operating at poor power

TABLE 5
List of Unsafe Control Actions for the Digital AVR, Operator and Management

Action By	Control Action	Not Providing Causes Hazard	Providing Causes Hazard	Too soon, Too late, Out of order	Stopped too soon, Applied too long
Digital AVR	Excitation Control	**UCA-AVR-1:** AVR does not provide excitation when coupled with the grid and operating at base load (generator operates as induction generator causing overheating of rotor) → [H-1.2, H-2.1] **UCA-AVR-2:** AVR does not increase excitation to match tie-line voltage preventing synchronization with the grid → [H-2.3] **UCA-AVR-3:** AVR does not increase excitation to achieve power factor or reactive power set-points, when synched with the grid, (utility financial penalties) → [H-2.1] **UCA-AVR-4:** AVR does not increase excitation when system voltage falls below lower voltage limit triggering system voltage collapse → [H-1.2] **UCA-AVR-12:** AVR does not increase excitation to required set-point when operating in islanded mode, causing undervoltage → [H-1.2]	**UCA-AVR-5:** AVR increases excitation when generator frequency decreases below synchronous speed leading to high V/Hz (overfluxing) during islanded or grid operation → [H-1.2] **UCA-AVR-6:** AVR increases excitation when generator terminal voltage is above setpoint (overvoltage) → [H-1.2] **UCA-AVR-7:** AVR increases excitation when turbine is tripped leading to high V/Hz (overfluxing) → [H-1.2] **UCA-AVR-11:** AVR increases excitation when a large inductive load is removed (overvoltage/overfluxing) → [H-1.2]	**UCA-AVR-8:** AVR increases excitation beyond ampere field no-load (AFNL) condition before generator is synchronized to the grid leading to high V/Hz (overfluxing) → [H-1.1.2; H-1.3]	**UCA-AVR-9:** AVR does not provide enough excitation (to maintain generator terminal voltage at grid voltage), when synched to the grid, resulting in pole slipping → [H-3; H-1.2] **UCA-AVR-10:** AVR increases excitation (too long) violating generator capability curve limits after synchronization with the grid (rotor overheating) → [H-1.2]
Plant Operator	Start/Stop Equipment Provide Setpoints	**UCA-OP-01:** Operator does not shut down equipment when hazardous conditions occur → [H-1]	**UCA-OP-02:** Operator provides incorrect setpoints to critical equipment during operation → [H-1, H-2]	**UCA-OP-02:** Operator shuts down critical equipment in the wrong order → [H-1, H-3]	
Plant Management	Issue Policies/ Funds	**UCA-PM-01:** Management does not disburse funds for cyber operations (e.g. patching) when vulnerabilities are identified → [H-1]	**UCA-PM-02:** Management approves insecure procedures to keep production running/during emergencies/contingencies → [H-1, H-2, H-3]	**UCA-PM-03:** Management disburses funding for training, mitigation strategy implementation too late → [H-1]	

factor is also penalized by the utility since it increases current flow through the distribution network (UCA-AVR-3).

If not corrected, the rotor field can weaken to the point that the gas turbine can cause the generator to 'slip a pole' i.e. generator rotor would suddenly turn as much as one complete revolution faster than it should be spinning and then violently come to a stop as it tries to magnetically link up again with the stator magnetic field. Such an event would cause catastrophic failure of the coupling between the turbine and the generator [41], [42] (UCA-AVR-9).

If there is a complete loss of excitation and the generator breaker is not tripped, it can cause the synchronous generator to operate as an induction generator, causing the rotor to quickly overheat, leading to insulation damage, high vibration, and rotor striking the stator, causing catastrophic damage [41], [42] (UCA-AVR-1). Apart from the generator, loss of excitation also impacts the power system; not only is a source of reactive power lost, but the plant acts as a reactive current sink to meet its reactive power demand which has the potential of triggering a system-wide voltage collapse [44] (UCA-AVR-4).

5.3.5 Under-voltage

When operating in *islanded* mode (i.e. independent of the grid), if the voltage drops too low, it has the potential to cause overheating of the motor loads at the plant due to an increase in current (to make up for the reduction in voltage), leading to overheating and pre-mature failure of motors [45] (UCA-AVR-12).

The systematic approach described above to identify UCAs for the AVR can be repeated for each of the controllers modeled in the functional control structure as documented by Khan [34].

5.4 Generate Loss Scenarios – Step 4

In the previous subsection, various system states were identified under which a given control action would be hazardous. In this section, we determine causal factors that enable the issuance of the earlier identified unsafe control actions.

According to *Leveson* [19], two types of causal scenarios must be considered (graphically shown in Fig. 7):

a. Scenarios that lead to the issuance of unsafe control actions; these could be a result of (1) *unsafe controller behavior* or (2) *inadequate/malformed feedback*.
b. Scenarios in which safe control actions are improperly executed or not executed altogether; these could be a result of issues along the (1) *control path* or the (1) *controlled process itself*.

For illustration purposes, we zoom into *the functional control structure* for the Digital AVR from Fig. 6 and superimpose it with guidewords from *Schmittner et al.* [32] signifying sample attack scenarios; the annotated control structure for the AVR is presented in Fig. 8. Using the *sample attack guidewords* as a starting point, scenarios are hypothesized which

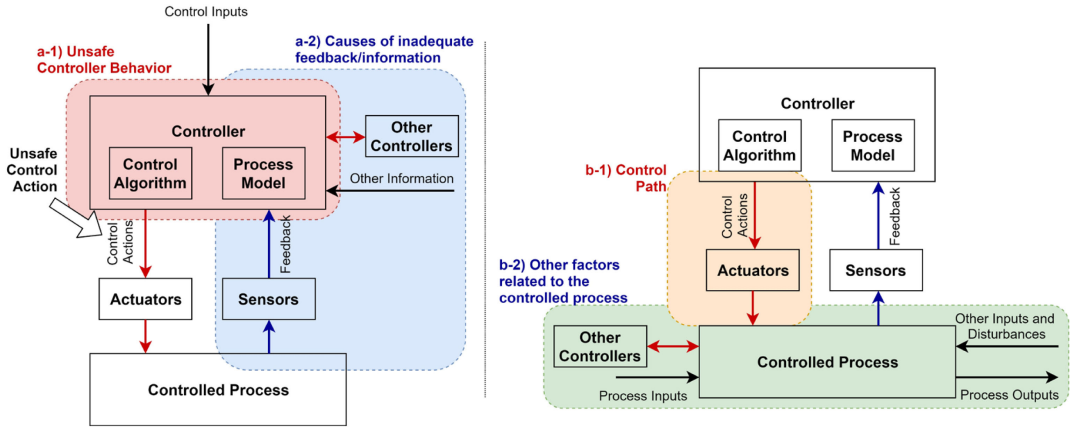

Fig. 7. Factors that can result in a) unsafe control actions b) safe control actions not or improperly executed.

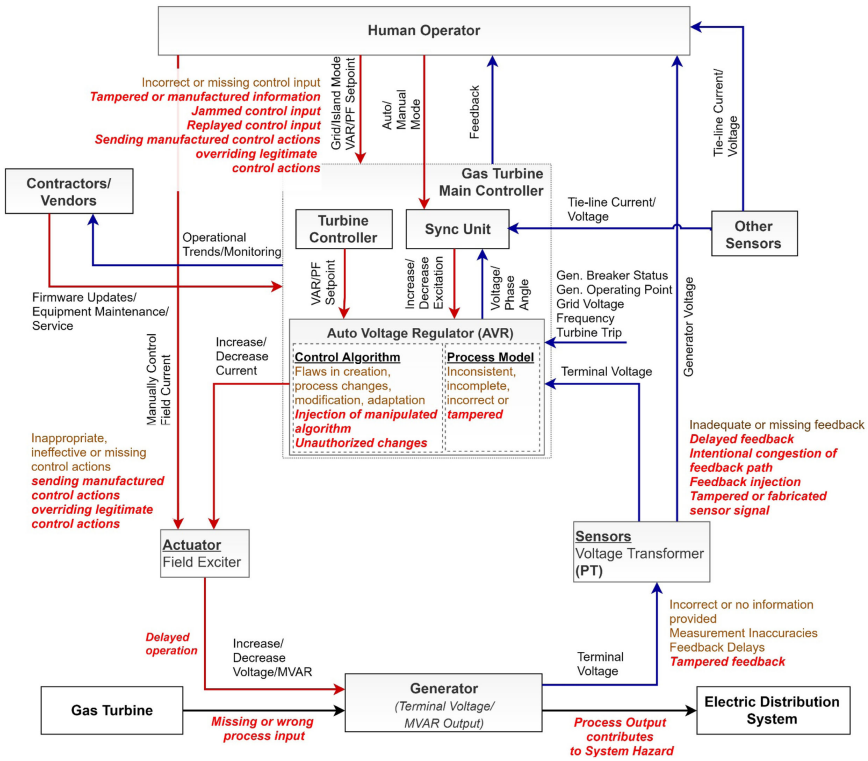

Fig. 8 Refined control-loop for the digital AVR.

would cause the controller to issue an unsafe command. For instance, the controller could issue an unsafe command because it is fed *tampered* data about the process state, or it could have the wrong process model to begin with (i.e. the process has changed over time, but the controller's process model has not been updated to reflect that change) etc.

Using this logic, each of the *sample attack guidewords* around the control loop are carefully contemplated as potential scenarios. Next, for each scenario, potential causal factors are determined. To generate insightful causal factors, we focus on six different categories of flaws for each scenario. These categories are:

- Process/Mental Model Flaws (assumptions)
- Contextual Factors (system or environmental)
- Structural Flaws (missing controls)
- Coordination & Communication Flaws
- Dynamics & Migration to Higher Risk
- Interdependencies Impact

Finally, *constraints* are then defined to prevent the issuance of the UCA as well as to mitigate the effects of the attack. Table 6 presents several *scenarios, associated causal factors* as well as *refined safety/security constraints* derived for the AVR control loop. A detailed discussion about the new insights gained by generating loss scenarios (in Table 6) is provided in the next section.

To further demonstrate the methodology, a few scenarios are generated for a *higher-level* controller as well i.e. the plant operator and presented in Table 7. As before, two type of scenarios are considered for the operator; either the operator provides an unsafe control input (i.e. pushing the wrong button) or does not adequately take the required

TABLE 6
List of Scenarios for the Digital AVR

UCA	AVR-08		
AVR increases excitation beyond ampere field no-load (AFNL) condition before generator is synchronized to the grid leading to high V/Hz (overfluxing) → [H-1.1.2; H-1.3]			
	Scenarios	**Associated Causal Factors**	**Safety/Security Constraints**
1	**Missing or incorrect feedback** AVR increases excitation beyond AFNL condition before generator is synced to the gird because of missing/incorrect feedback from PTs. This could be caused by: a) Network-based malicious spoofing of PT values by stealthy manipulation e.g. the attacker alters AVR I/O register values that store the current voltage state b) Insider attack – attacker removes the PT fuse or cuts the sensor (PT) wire so that AVR does not receive any feedback; loss of voltage sensing protection does not act to shut down the generator	**1. Process Model Flaws/Assumptions** • AVR intrinsically trusts the sensor values – it believes field current needs to be increased to meet excitation setpoint but does not verify if the sensor value is correct **2. Contextual Factors** • AVR is connected to plant control system via Modbus protocol i.e. after the attacker gains access to the OT network, the attacker can modify AVR registers/coils via network-based attacks **3. Structural Flaws** • Missing high V/hz (overfluxing) protection • Poor physical access controls; insider/attacker has physical access to AVR control cabinet (i.e. could remove a fuse or strip a wire to disable voltage sensing function) • No intrusion detection and monitoring at Level 0	1. The side must employ Endpoint Detection & Response (EDR) solution 2. Overflux relay ANSI 24 must be installed **with separate PT** 3. AVR must alarm the operator if feedback is not received 4. Physical access control must be bolstered; access to control cabinets must be restricted on a need-to-access basis with adequate security controls.
2	**Malformed process model as a result of tampered feedback** AVR increases excitation beyond AFNL condition before generator is synced to the gird because of a malformed process model; incorrect feedback about generator breaker status	**1. Process Model Flaws/Assumptions** • AVR believes synchronization has already occurred and increases reference excitation signal to attain VAR/PF setpoint when generator breaker is still open **2. Contextual Factors** • AVR is connected to plant control system via Modbus protocol **3. Coordination & Communication Flaw** • AVR receives the 'Breaker Closed' signal before the breaker is actually closed due to malicious alteration of generator breaker *coil value*	1. AVR must have physical interlock with generator breaker 2. Overflux Relay ANSI 24 must be installed - *Same as SC-AVR-08-01-03* 3. Overvoltage relay ANSI 59 must be installed ***(Already installed)***
3	**Unauthorized changes to control algorithm** AVR increases excitation beyond AFNL condition before generator is synced to the gird because of unauthorized changes to its control algorithm i.e. changes to control logic/setpoints leading to incorrect sequence/excessive voltage at startup. This could be caused by a: a) Contractor via remote access* b) Malicious insider c) Cyber attacker with access to OT network	**1. Process Model Flaws/Assumptions** • Belief that contractor has been vetted and follows same rigorous security procedures as the asset owner **2. Contextual Factors** • Engineering workstation has access to source code for all programs running at site (more than what is needed) • AVR is connected to engineering workstation to enable diagnostics and programming • Contractors/3rd parties have remote access to engineering workstation **3. Structural Flaws** • Inadequate change management procedures for AVR source code in place (i.e. anyone with access to the engineering workstation can upload a program to the AVR); engineering workstation has poor access control	1. The site must ensure that access to source code/programs running on field devices is severely restricted and follows the *least privilege* information security principle 2. Field device/PLC source code must be secured 3. The site must employ validation using digital signatures or checksums to ensure integrity of code running on field devices at a regular frequency 4. Contractor access including remote access protocols must be analyzed
	* **Note:** Only causal factors related to ***contractor*** actions are discussed here due to space limitations; the same approach could be extended for other categories	• No controls in place to observe contractor actions via remote access **4. Dynamics & Migration to Higher Risk** • AVR has physical security features to prevent remote programming. However, to facilitate convenience, the plant enabled remote programming functionality on the AVR	5. Remote programming feature of AVR must be disabled
4	**Inadequate process model** AVR increases excitation beyond AFNL condition before generator is synced to the gird because of an inadequate process model – a spurious loss of feedback condition causes a bumpless transfer to manual mode which is exploited by an attacker	**1. Process Model Flaws/Assumptions** • Assumption that the loss of feedback is due to a safety issue when in reality it is caused by an attacker **2. Contextual Factors** • AVR is connected to plant control system via Modbus; an attacker can send manufactured feedback to AVR to change its state • AVR is coded with bumpless transfer scheme to enable operator to take control when AVR fails to raise excitation to the required level	1. A physical interlock with generator breaker for AVR loss of feedback condition must be implemented
Impact on System Mission and/or Interdependent Systems If an overfluxing condition is not arrested within a few seconds, it would destroy the generator windings beyond repair, preventing on-site generation. This would result in the plant drawing more power from the grid to meet its demand, increasing operational cost. Loss of generator would logically imply loss of turbine as a heat source for the Heat Recovery Steam Generator, which would imply downgraded steam production which in turn would impact chilled water production since some of the chillers are steam-driven. This would likely trigger an increase in output for the electric-driven chillers which would further increase electricity drawn from the grid.			

corrective action (i.e. the corrective action is rendered ineffective).

By following a similar approach, we can move around the control structure presented in Fig. 6 and evaluate each of the other controllers in order to generate a complete set of causal factors that lead to system-level losses.

6 Discussion

The causal factors identified in Step 4 above (presented in Tables 6 and 7) are indicative of different types of vulnerabilities. These include vulnerabilities that are local to the control loop being analyzed (component failure flaws), vulnerabilities emerging as a result of interactions between components (based on interdependencies external to the control loop) and unsafe control inputs from hierarchical controllers (or the violation of constraints due to ineffective implementation of controls by higher-level controllers). Note that any of these vulnerabilities may be exploited by a malicious actor.

6.1 Types of Vulnerabilities Discovered

For instance, the analysis revealed that in the event that the AVR malfunctions and causes the generator to overflux, the

TABLE 7
List of Scenarios for the Operator

UCA	OP-02		
Operator provides incorrect setpoints to equipment during operation (leading to equipment damage) → [H-1; H-2]			
	Scenarios	**Associated Causal Factors**	**Safety/Security Constraints**
1	**Incorrect feedback** DCS/HMI displays incorrect process feedback (voltage) because attacker uses Complex Measurement Response Injection (CMRI) technique to update feedback signal from PLC. The feedback injection causes operator to have a malformed process model (through the DCS/HMI) who then takes over manual control and provides incorrect control settings (e.g. increasing excitation beyond AFNL condition)	**1. Process Model Flaws/Assumptions** • Belief that HMI is providing correct information - prior to taking over manual control operator does not independently verify voltage **2. Contextual Factors** • Operators are not exposed to such attack vectors (lack of cybersecurity training) • Organizational pressure to keep the plant running **3. Structural Flaws** • Field device i.e. AVR protections (over voltage, under voltage, high V/Hz etc.) are improperly set and do not prevent the attack **4. Dynamics & Migration to Higher Risk** • Plant resorted to using default passwords on field devices (e.g. AVR) since password management process is not set up	1. Operator must have independent out-of-band feedback about generator voltage level 2. Overflux Relay ANSI 24 must be installed – same as *SC-AVR-08-01-03* 3. Management must ensure adequate training is provided to escalate spurious behavior of equipment and improve cyber-awareness and security culture 4. Management must ensure a robust password management policy
2	**Sending manufactured control actions** Attacker issues incorrect settings from DCS masqueraded as coming from the operator; attacker overrides legitimate control actions by employing a Malicious Parameter Command Injection (MPCI) attack, modifying control parameters on field devices (such as the AVR) while jamming operator commands	**1. Process Model Flaws/Assumptions** • Belief that a command originating from the HMI is provided by the operator and hence legitimate when it could have been provided by the attacker **2. Contextual Factors** • Field device mode (auto vs. manual) can be remotely changed – allowing attacker to take over control of the field device **3. Structural Flaws** • Operator does not have out-of-band control of field device (AVR) such that control of the compromised device could be taken over in case malicious behavior is observed	1. Unauthorized access to the DCS/Turbine Controller/AVR must be prevented 2. Out-of-band control and feedback must be provided to the operator - Same as *SC-OP-02-1-01* 3. AVR must have a physical switch to convert between auto and manual modes
3	**Injection of manipulated control algorithm** Operator knows the actual state of the generator's terminal voltage and the status of the generator breaker yet provides the wrong control input to the AVR. Critical parameters (such as AFNL condition set-point) is maliciously altered in the operator's procedure	**1. Process Model Flaws/Assumptions** • Belief that the operator procedure is correct and up to date **2. Contextual Factors** • Procedure is stored on an unsecured part of the network **3. Structural Flaws** • No controls in place to ensure integrity of operator procedures • No controls to secure operator procedures **4. Dynamics & Migration to Higher Risk** • Process to escalate errors in operating procedures is convoluted and bureaucratic. It is generally accepted that procedures are correct; feedback between engineers and operators is non-existent	1. Operator must have hardcopy of quality controlled-procedure in the control room 2. Management must ensure operator is provided adequate training 3. Management must ensure technical specs., operating procedures are stored securely and develop processes to validate integrity 4. Operating procedure error reporting must be streamlined and enforced

protection scheme at the plant is not equipped with an overflux relay (ANSI Device Code 24) to control such a situation – an example of a component-level flaw that was discovered through the analysis. The fact that the protection scheme is missing an overflux relay is not completely unexpected; *Scharlach* [43] notes that traditionally the overflux protection is implemented in generators larger than 100 MW but cautions that "*due to the serious effects that can result from an undetected overexcitation event*", this protective element should be applied even on smaller machines, especially given the low cost of such relays.

Note that the overflux relay is required in the event that the AVR fails to enforce the required constraints on the generator terminal voltage. We will now explore the scenarios that would cause it to violate its safety and security constraints in the first place. Scenario SC-AVR-08-01 (Scenario 1 in Table 6) is rather intuitive – if the AVR is provided incorrect information about the state of the terminal voltage, it would produce incorrect voltage outputs resulting in a loss – it is doing what it is designed to do. In addition, the associated causal factors provide interesting insights about the flaws in the system that would enable such a loss scenario to succeed. For instance, there is implicit *trust without verification* in the interaction between the sensor and the controller. Coupled with the use of insecure industrial protocols, missing overfluxing protection and poor physical access controls, we can see how this is really a system problem that needs a system level solution.

Scenario SC-AVR-08-02 is similar but involves feedback about a component state that is not part of the AVR control loop – i.e. informational interdependency external to the control loop. Here, the assumption is that the generator breaker is closed when in reality it could simply have been spoofed by an attacker. This raises questions about the design of the process (in terms of information exchange integrity between components) and additional controls that could prevent such an outcome. This scenario is an example of a loss scenario resulting from interaction between components (the generator breaker and the AVR).

Scenario SC-AVR-08-03 is different from the other ones presented thus far because it involves a change in the control algorithm of the AVR controller. It shows how access to the AVR by an external contractor for legitimate business purposes could be exploited by an adversary resulting in a loss. The associated causal factors highlight flaws in assumptions and controls including beliefs about the vetting process of contractors and 3rd parties, their level of access to critical assets (such as engineering workstations) and observeability of their actions in making changes to sensitive equipment. This scenario also highlights the tradeoff between functionality and security; remote access and control enables convenience, but also increases the attack surface. It also highlights the weakeness in process of storing more information that what is necessary on the engineering workstation and providing access to an unvetted 3rd party via remote access.

Scenario SC-AVR-08-04 highlights an important assumption that can be exploited by an adversary; the design of the process does not account for malicious intent. A spurious loss of feedback condition from the AVR, by design, transfers the operation mode from automatic to manual which can then be exploited by an attacker to manually increase voltage of the generator beyond safe limits.

Finally, the last row in Table 6 hypothesizes the system impact of the UCA i.e. the ability of the AVR to produce effects that go beyond the AVR control loop. For instance, hazardous function of the AVR would cause damage to the generator which would impact the aggregate steam output from the heating system, which in turn would require the steam-driven chillers to be shut-down in preference of the electric-driven chillers, which in turn would additionally stress the electric distribution system because of additional import from the grid. This again highlights the need for taking a systems perspective of the security problem.

Table 7 lists three scenarios for the plant operator where the operator either takes the wrong action (because of bad information or otherwise) or provides inadequate corrective action. Similar to the AVR's process model, the operator also has a mental model of the various processes in the plant, albeit at a higher level of abstraction. Scenario SC-OP-02-01 shows how in the absence of out-of-band feedback, the operator may be convinced of AVR malfunction through malicious feedback injection, forcing the operator to manually override the AVR and increase excitation resulting in the loss. One important point to note here is that this scenario is possible because of poor cybersecurity culture and training etc., that may emerge as a result of management's focus *'on keeping the plant running'* – putting pressure on the operator to try to resolve the issue without escalating it to engineering/management.

Scenario SC-OP-02-02 describes how the ability to remotely alter field device mode (from *auto* to *manual*) can be exploited by an adversary to take over control of the physical process to launch an attack.

Scenario SC-OP-02-03 describes how the operator's control algorithm may be compromised by altering voltage setpoints in the operating procedure – an indirect effect of not having access to physical copies of operating procedures in the control room. The causal factors for this scenario also highlight cultural aspects; lack of direct engagement between engineers and operators has made the process of reporting errors convulted and difficult. This can result in a scenario where the operator follows the procedure at face-value even if the prescibred settings are incorrect.

6.2 Mitigation Requirements

Note that despite the limited nature of the analysis, different types of vulnerabilities have been uncovered. Fig. 9 illustrates some high-level functional requirements and changes to the design of the control structure that could prevent or mitigate the effects of an attack on the AVR. For instance, the functional requirements presented in Fig. 9 recommend addition of an out-of-band feedback loop for generator terminal voltage to the operator as well as the implementation of an overflux relay into the protection scheme. In addition, some design changes are recommended as functional requirements such as interlocking generator voltage with generator breaker and generator frequency to preclude an overfluxing event altogether i.e. to eliminate the vulnerability through design, if possible.

Furthermore, process changes are suggested in the form of changes to the operator procedure when synchronizing the generator to the grid. New requirements also include changes to engineering responsibilities in terms of access and storage of operating procedures in the control room. Finally, additional constraints are recommended on management as well as outside vendors and contractors in the form of policy changes for access to plant equipment. The important thing to note is that these functional requirements span all levels of the control structure – technical, process, human and organizational.

6.3 Cybersafety Performance Evaluation

As noted earlier, *Cybersafety* is a strictly qualitative method. In this subsection we provide a subjective evaluation of *Cybersafety's* performance in identifying and mitigating cyber-vulnerabilities.

Step 1 (define basis of the analysis) was the easiest to implement but required significant input from the key stakeholders (operators, engineers as well as management) to ensure that the target system was correctly identified and that critical losses and hazards were correctly defined. From a duration perspective, this step required an engagement of a few hours with the key stakeholders.

Step 2 (model the functional control structure) required several iterations to get right. It also required input from all the key stakeholders. This step was quite insightful for all parties involved as it helped to clarify the control structure within the organization. It was interesting to observe that there was divergence in understanding about how the various components of the system interacted with one another. Agreeing upon a functional control structure, was in of itself, very valuable to the organization, since it enhanced clarity of function for the various controllers in the organization. From a duration perspective, this step took a few hours to complete (the actual duration was longer, as it was iterated several times).

Step 3 (identify unsafe control actions) was straight forward but required individual engagements with subject-matter experts (SME). It was observed that for the most part, the SMEs were able to list UCAs fairly easily without actually developing detailed process models. However, identifying UCAs for higher-level controllers was found to be rather abstract and subjective. From a duration perspective, this step took several days to complete. It should be noted that this step is dependent upon the level of abstraction of the system. At the level of abstraction shown in this study we elicited approximately 90 UCAs in the course of 4-5 days.

Step 4 (generate loss scenarios) was by far the most challenging to complete. There was very little guidance in the publicly available STAMP/STPA literature [19] for this phase of execution. However, by using the sample attack guidewords and thinking about the causal factors in terms of flaw categories and missing controls, we were able to generate rich causal factors that were able to capture the dynamics of the system and indirect interactions. We

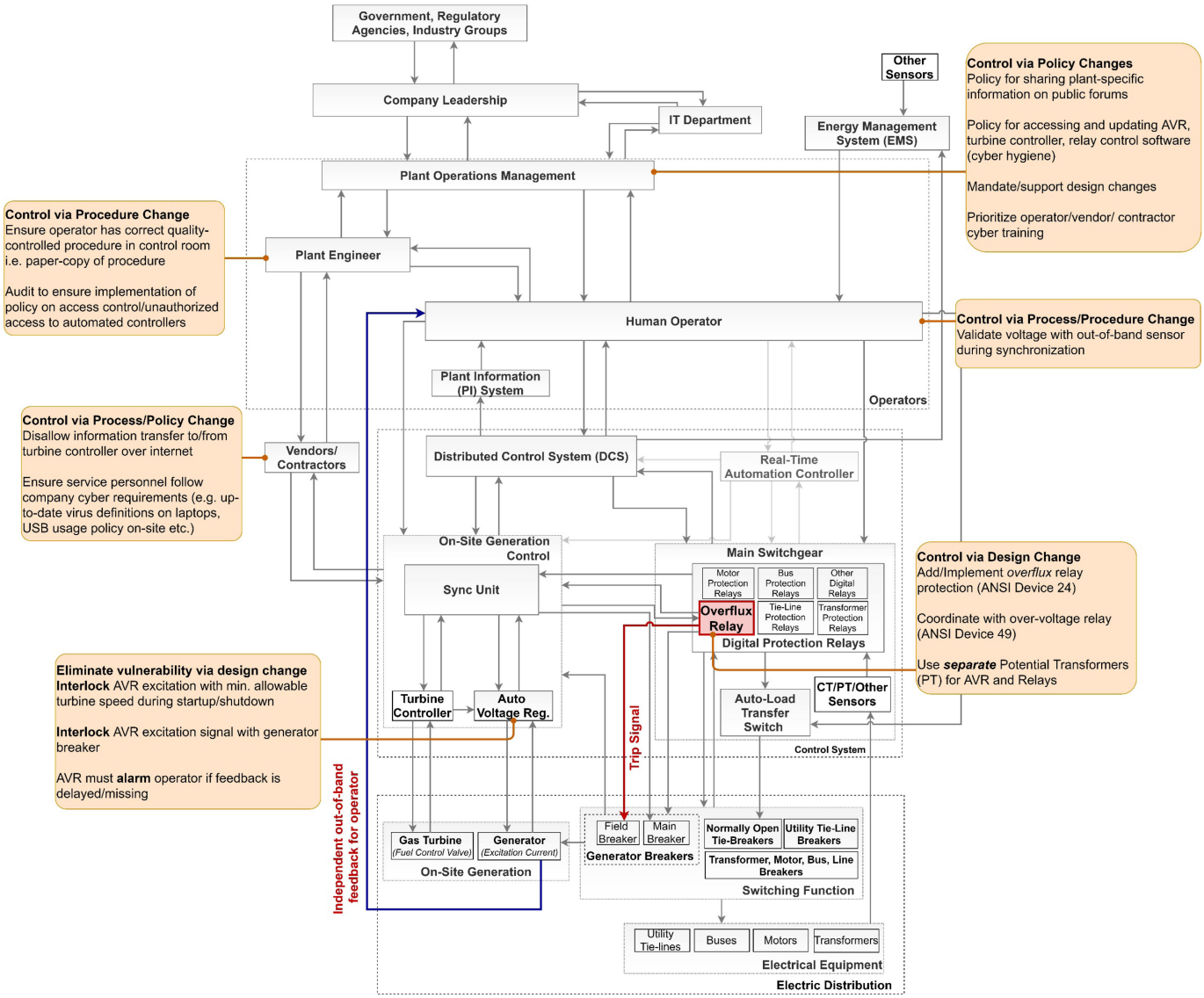

Fig. 9. A subset of proposed requirements/constraints to eliminate and/or mitigate vulnerabilities.

generated scenarios and causal factors for only 10 percent of the UCAs which took 2-3 weeks to complete. This step required significant subject-matter knowledge but was a creative exercise that led to insightful discussions with the stakeholders about controls, threats and vulnerabilities.

Overall, each phase of execution highlighted different aspects of the system's design and its weakenesses which deepened understanding about the system, its vulnerabilities and attack surface.

7 Conclusions

The objective of this study was to demonstrate the ability of the *Cybersafety* method to systematically and robustly uncover cyber vulnerabilities and mitigation strategies in an industrial control system using a real-world example; specifically, those vulnerabilities that emerge as a result of interactions between components and interdependent subsystems.

We demonstrated the application of *cybersafety* to identify cyber-vulnerabilities in an archetypal industrial control system. This was a first-of-a-kind analysis on the cyber vulnerabilities of the electric generation and distribution system using a systems perspective. It was discovered that the addition of a few additional steps, makes the method more robust and repeatable and makes the analysis more comprehensive. The effect of system interdependencies was included in the analysis which influenced each step of the analysis; from the problem statement in Step 1, to the modeling of the extended functional control structure in Step 2, to the identification of process model variables in Step 3 and finally generation of loss scenarios and impact on the system in Step 4.

Using the analogy of the human body, just as it is impossible to avoid all contact with infections and never catch a disease, it is impossible for an industrial control system to be under constant attack and never have its network defenses breached. Therefore, the system has to be designed so that it is resilient against the effects of the attack and *Cybersafety* provides a well-guided and structured analytical method to identify vulnerabilities and derive functional requirements to improve resilience against cyberattacks.

ACKNOWLEDGMENTS

This work was supported by Department of Energy under Award DE-OE0000780, in part by MIT Energy Initiative (MITei), and in part by the corporate members of Cybersecurity at MIT Sloan: The Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity.

REFERENCES

[1] G. Loukas, *Cyber-Physical Attacks : A Growing Invisible Threat*. Cambridge, U.K.: Butterworth-Heinemann, 2015.

[2] M. G. Angle, S. Madnick, J. L. Kirtley, and S. Khan, "Identifying and anticipating cyber attacks that could cause physical damage to industrial control systems," Accessed: May 4, 2019. [Online]. Available: http://web.mit.edu/smadnick/www/wp/2017-14.pdf

[3] S. G. Freeman, C. St Michel, R. Smith, and M. Assante, "Consequence-driven cyber-informed engineering (CCE)," Idaho Falls, ID, USA, 2016. [Online]. Available: https://doi.org/10.2172/1341416

[4] GE Grid Solutions, "STV overexcitation relay." Accessed: Dec. 19, 2019. [Online]. Available: https://www.gegridsolutions.com/multilin/catalog/stv.htm

[5] "General electric LM2500 | powerweb," Accessed: Dec. 19, 2019. [Online]. Available: http://www.fi-powerweb.com/Engine/Industrial/GE-LM2500.html

[6] A. A. Cárdenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," Accessed: May 1, 2019. [Online]. Available: https://ptolemy.berkeley.edu/projects/chess/pubs/601/cps-security-challenges.pdf

[7] C. Johnson, "Why we cannot (yet) ensure the cyber-security of safety-critical systems," in *Proc. 24th Safety-Crit. Syst. Symp.*, 2016, pp. 171–182.

[8] C. Schmittner, T. Gruber, P. Puschner, and E. Schoitsch, "Security application of failure mode and effect analysis (FMEA)," in *Proc. Int Conf. Comput. Safety Rel. Secur.*, 2014, pp. 310–325

[9] H. A. Duckworth and R. A. Moore, *Social Responsibility: Failure Mode Effects And Analysis*. Hoboken, NJ, USA: CRC Press, 2010.

[10] M. Steiner and P. Liggesmeyer, "Combination of safety and security analysis - finding security problems that threaten the safety of a system," in *Proc. 32nd Int. Conf. Comput. Safety Rel. Secur.*, 2013, pp. 1–8

[11] H.A. Watson, *Launch Control Safety Study*. Murray Hill, NJ, USA: Bell Labs, 1961.

[12] A. Altawairqi and M. Maarek, "Attack modeling for system security analysis," in *Proc. Int. Conf. Comput. Safety Rel. Secur.*, 2017, pp. 81–86.

[13] C. A. Ericson, *Hazard Analysis Techniques for System Safety.*, Hoboken, NJ, USA: Wiley, 2005.

[14] L. Dawson, A. Muna, T. Wheeler, P. Turner, G. Wyss, and M. Gibson, "IAEA-CN-228-12 1 assessment of the utility and efficacy of hazard analysis methods for the prioritization of critical digital assets for nuclear power cyber security," Accessed: May 1, 2019. [Online]. Available: www.osti.gov/servlets/purl/1252915

[15] X. Xu, M. L. Ulrey, J. A. Brown, J. Mast, and M. B. Lapis, "Safety sufficiency for NextGen assessment of selected existing safety methods, tools, processes, and regulations," Accessed: May 2, 2019. [Online]. Available: http://www.sti.nasa.gov

[16] N. G. Leveson, "Is estimating probabilities the right goal for system safety?," Accessed: Oct. 7, 2019. [Online]. Available: http://sunnyday.mit.edu/papers/Making-Safety-Decisions.pdf

[17] J. Dunjó, V. Fthenakis, J. A. Vílchez, and J. Arnaldos, "Hazard and operability (HAZOP) analysis. A literature review," *J. Hazardous Mater.*, vol. 173, pp. 19–32, 2010.

[18] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, and S. Sezer, "STPA-SafeSec: Safety and security analysis for cyber-physical systems," *J. Inf. Secur. Appl.*, vol. 34, pp. 183–196, 2016.

[19] N. G. Leveson and J. P. Thomas, "STPA handbook," Accessed: Apr. 28, 2019. [Online]. Available: http://psas.scripts.mit.edu/home/

[20] N. Leveson, *Engineering A Safer World : Systems Thinking Applied To Safety*. Cambridge, MA, USA: MIT Press, 2012.

[21] J. P. Thomas, IV, "Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis," Ph.D. dissertation, Eng. Syst. Division, Massachusetts Inst. Technol., Cambridge, MA, USA, 2013.

[22] R. S. Martínez, "System theoretic process analysis of electric power steering for automotive applications," M.Sc. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2015.

[23] N. A. Peper, "Systems thinking applied to automation and workplace safety signature of author," M.Sc. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2007.

[24] J. Thomas, F. L. De Lemos, and N. Leveson, "Evaluating the safety of digital instrumentation and control systems in nuclear power plants," Accessed: May 2, 2019. [Online]. Available: http://sunnyday.mit.edu/papers/MIT-Research-Report-NRC-7-28.pdf

[25] P. D. Stukus, "Systems-theoretic accident model and processes (STAMP) applied to a U.S. coast guard buoy tender integrated control system," M.Sc. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2017.

[26] T. Pawlicki, A. Samost, D. W. Brown, R. P. Manger, G.-Y. Kim, and N. G. Leveson, "Application of systems and control theory-based hazard analysis to radiation oncology," *Med. Phys.*, vol. 43, pp. 1514–1530, 2016.

[27] J. R. Laracy, "A systems-theoretic security model for large scale, complex systems applied to the US Air Transportation System," Accessed: May 2, 2019. [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/39256/173417210-MIT.pdf?sequence=2&isAllowed=y

[28] J. R. Laracy, "Applying STAMP to critical infrastructure protection," Accessed: May 2, 2019. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.7643&rep=rep1&type=pdf

[29] H. M. Salim, "Cyber safety: A systems thinking and systems theory approach to managing cyber security risks," Accessed: May 21, 2019. [Online]. Available: https://dspace.mit.edu/handle/1721.1/90804

[30] A. Nourian and S. Madnick, "A systems theoretic approach to the security threats in cyber physical systems applied to stuxnet," *IEEE Trans. Dependable Secur. Comput.*, vol. 15, no. 1, pp. 2–13, Jan./Feb. 2018.

[31] W. Young and N. G. Leveson, "An integrated approach to safety and security based on systems theory," *Commun. ACM*, vol. 57, pp. 31–35, 2014.

[32] C. Schmittner, Z. Ma, and P. Puschner, "Limitation and improvement of STPA-sec for safety and security co-analysis," in *Proc. Int. Conf. Comput. Safety Rel. Secur.*, 2016, pp. 195–209.

[33] 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.

[34] S. Khan, "Using a system-theoretic approach to identify cyber-vulnerabilities and mitigations in industrial control systems," M. Sc. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2019.

[35] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," Accessed: Apr. 27, 2019. [Online]. Available: https://pdfs.semanticscholar.org/b1b7/d1e0bb39badc3592373427840a4039d9717d.pdf

[36] J. Thomas, "Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis," 2012. [Online]. Available: https://doi.org/10.2172/1044959

[37] "Guard against over-fluxing: Ensure proper generator protection, maintenance," Accessed: Apr. 23, 2019. [Online]. Available: http://www.ccj-online.com/guard-against-over-fluxing-ensure-proper-generator-protection-maintenance/

[38] VOITH, "ThyriconTM excitation system," Accessed: Apr. 23, 2019. [Online]. Available: https://voith.com/cn/t3387_e_Thyricon_screen.pdf

[39] GE Automation & Controls, "EX2100e excitation control 100 mm, 77 mm, 53 mm, and 42 mm thyristor systems product description," Accessed: Apr. 23, 2019. [Online]. Available: https://www.ge.com/content/dam/gepower-new/global/en_US/downloads/gas-new-site/resources/automation/gea-s1302_ex2100e_100mm_77mm_53mm_42mm_thyristor_systems.pdf

[40] IEEE IPES Generator Protection Task Force, "IEEE tutorial on the protection of synchronous generators," Accessed: Apr. 24, 2019. [Online]. Available: https://www.pes-psrc.org/kb/published/reports/IEEEGenProtTutorial_20110506_1file.pdf

[41] C. J. Mozina, "Power plant "horror stories"," in *Proc . IEEE Power Eng. Soc. Inaugural Conf. Expo. Afr .*, 2005, pp. 462–465.

[42] G. Klempner and I. Kerszenbaum, *Operation and Maintenance of Large Turbo-Generators*. Hoboken, NJ, USA: Wiley, 2004.
[43] R. C. Scharlach and J. Young, "Lessons learned from generator event reports," Accessed: Apr. 24, 2019. [Online]. Available: https://cms-cdn.selinc.com/assets/Literature/Publications/Technical%20Papers/6387_LessonsLearnedGeneratorRS-JY_20100303_Web2.pdf?v=20191007-202032#:~:text=Additionally%2C%20the%20effects%20of%20negative,each%20function%20to%20protect%20generators
[44] W. Hartmann, "Generator protection overview," Accessed: Apr. 25, 2019. [Online]. Available: https://www.eiseverywhere.com/file_uploads/8b5452d7f9376912edcba156cd1a5112_WSU_GENPROTOVERVIEW_180305.pdf
[45] UST Power, "AVR guide: Voltage too high, too low | UST," Accessed: Apr. 24, 2019. [Online]. Available: https://ustpower.com/comparing-automatic-voltage-regulation-technologies/need/avr-guide-voltage-high-low/

Shaharyar Khan received the BASc (Hons.) degree in mechanical engineering from the University of Waterloo in 2010 and the SM degree in engineering and management from the Massachusetts Institute of Technology in 2019. He is currently a research scientist with the MIT Sloan School of Management. He was a seismic/structural design engineer with BWX Technologies, designing and analyzing critical components for nuclear power plants. He was also the lead project engineer with Nuclear Generating Station, deploying tools for reactor inspections and maintenance. He is a registered professional engineer, Ontario, Canada.

Stuart Madnick (Member, IEEE) received the SB degree in electrical engineering, the SM degreee in management, and the PhD degree in computer science from MIT. He was the John Norris Maguire professor of information technology and a professor of engineering systems in 1960, and since 1972, he has been an MIT faculty member. He was the head with Information Technologies Group, Sloan School of Management, MIT, for more than 20 years. He is the founding director of Cybersecurity at MIT Sloan (CAMS), the Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity. He is the author or coauthor of more than 350 books, articles, or reports, including the classic textbook on operating systems, and has three patents. His current research interests include cybersecurity, information integration technologies, semantic web, database technology, software project management, and the strategic use of information technology. He has been active in industry, as a key designer and developer of projects, such as IBM's VM/ 370 operating system and Lockheed's DIALOG information retrieval system. He was a consultant to major corporations and has been the founder or cofounder of five high-tech firms, and currently operates a hotel in the 14th century Langley Castle in U.K.