Cybersafety:
A System-theoretic Approach to Identify Cyber-vulnerabilities & Mitigations in Industrial Control Systems

Shaharyar Khan, Stuart Madnick

Working Paper CISL# 2019-22

December 2019

Cybersecurity Interdisciplinary Systems Laboratory (CISL)
Sloan School of Management, Room E62-422
Massachusetts Institute of Technology
Cambridge, MA 02142

Cybersafety: A System-theoretic Approach to Identify Cyber-vulnerabilities & Mitigations in Industrial Control Systems

Shaharyar Khan, Stuart Madnick, Member, IEEE

Abstract— Recent cyber-physical attacks, such as Stuxnet, Triton etc., have invoked an ominous realization about the lethality of such attacks and the vulnerability of critical infrastructure, including power, gas and water distribution systems. Traditional IT security-biased protection methods that narrowly 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 systems in an integrated, holistic fashion that leverages the physics of the cyber-physical system. System-Theoretic Accident Model & Processes (STAMP) offers a powerful, holistic, structured 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 an actual electrical generation and distribution system of an archetypal industrial facility to demonstrate the application of a STAMP-based method – we call Cybersafety – to identify and mitigate cyber-related 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 robust and structured format such that it can be emulated to analyze 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 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 mitigations is provided in Section 6 followed by a short conclusion in Section 7.

- 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).

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 intensions 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 the safety and security worlds and provision of guide words to elicit scenarios due to malicious actions in the final step of the analysis.

Similar to STPA-Sec, *Friedberg et al.* [18] present an analysis 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.

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.

Despite best effort, our literature search did not reveal any detailed published work documenting the application of STPA-Sec to industrial control systems or power generation plants. 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. This *focused* approach to identify and mitigate cyber-related vulnerabilities in ICS is called *Cybersafety* and is described in the following section.

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 emerging from controller interactions.

3 CYBERSAFETY

The basic steps in the *Cybersafety* method are identical to STPA and summarized in Figure 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 malicious 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

Finally, new functional requirements and mitigation strategies are defined that would prevent the worst-possible outcomes identified in Step 1.

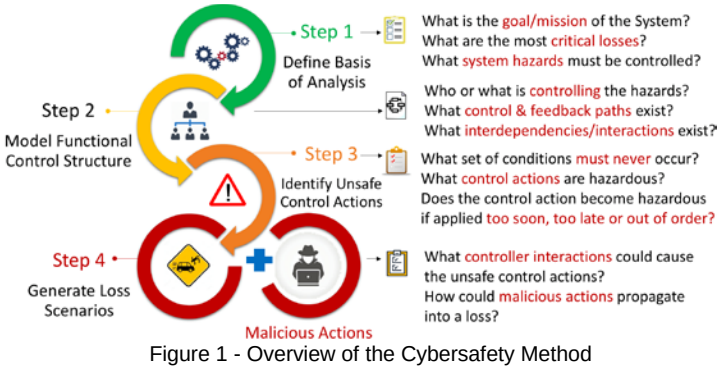

Figure 1 - Overview of the Cybersafety Method

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.

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 Siemens 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 Figure 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* [33].

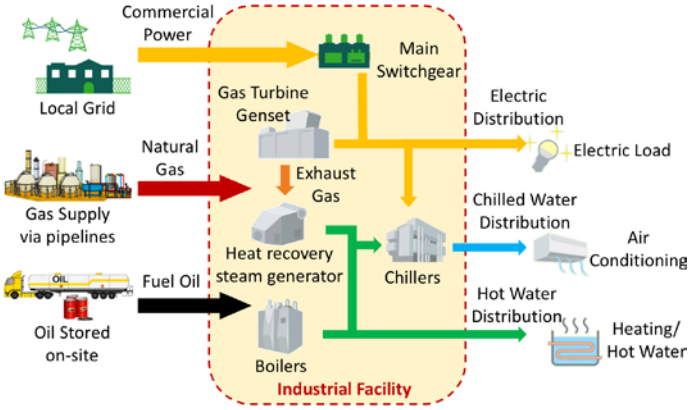

Figure 2 - The Plant - A Microcosm Energy Facility

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 [33]. 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 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 Figure 3.

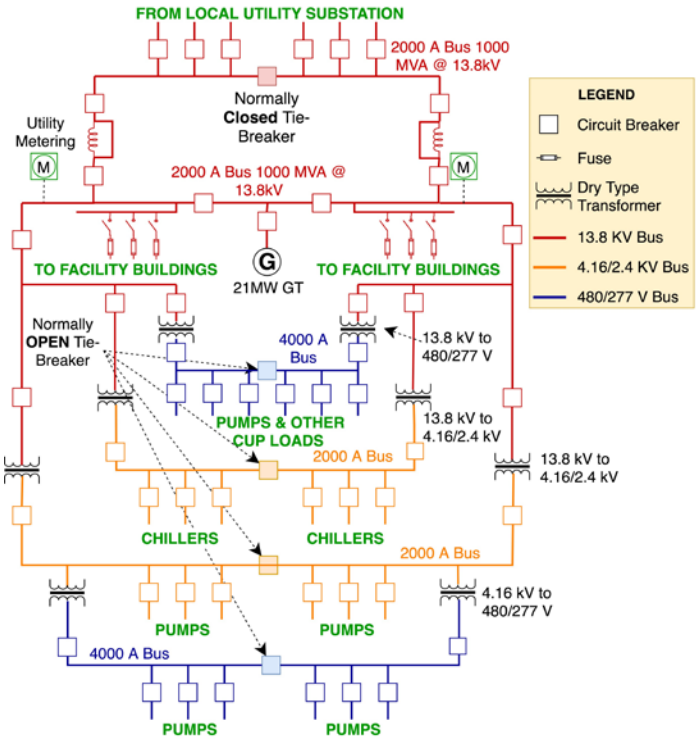

Figure 3 - One-line Diagram of Plant's Electrical Distribtion System

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 set-points 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 critical functions of the system as shown in Figure 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.

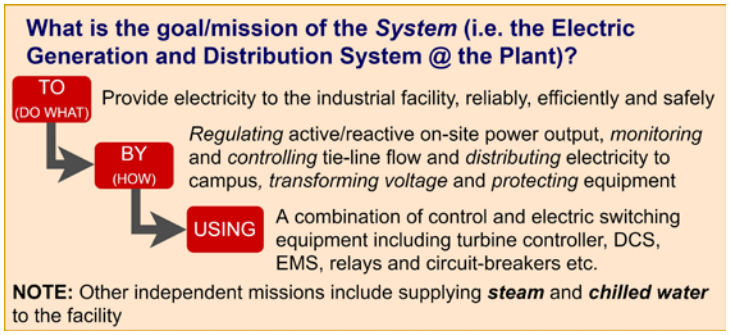

Figure 4 - System Problem Statement

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 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).

Table 1 - List of Unacceptable High-Level System Losses

	System-Level Losses
L-1	Death, dismemberment or injury to personnel
L-2	Physical damage to critical equipment
L-3	Loss of mission i.e. inability to deliver electricity to campus
L-4	Economic loss due to an electrical event (including capital cost or operational cost)

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 were found to be in vain; either the hazards were too high-level to provide any meaningful information about the system or they were defined in terms of system 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 which can inform the development of the *functional control structure* in Step 2 of the method. The coarse list of hazards is provided in Table 2, while the refined list of hazards is provided in Table 3.

Table 2 - Coarse list of System-Level Hazards

	System-Level Hazards
H-1	Inadequate regulation/control of on-site generation
H-2	Inadequate monitoring and control of supply tie-line
H-3	Inadequate isolation of faulted equipment
H-4	Inadequate voltage transformation

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 4. Progressing through the analysis, this list of constraints is systematically refined, ultimately, resulting in a set of functional requirements to protect against specific hypothetical loss scenarios in Step 4.

Table 3 - Refined list of System-Level Hazards

System-Level Hazards	Related Losses
H-1: Inadequate regulation/control of on-site generation	
H-1.1: Equipment is operated beyond normal operational limits H-1.1.1: Mechanical parameters (speed, ramp rates, temp., pressure, vibration as applicable) exceed normal operational limits H-1.1.2: Electrical parameters (current, voltage, frequency) exceed normal operational limits	L-1, L-2, L-3, L-4
H-1.2: Active and reactive power is not adequately controlled H-1.2.1: Violation of power quality metrics H-1.2.2: On-site generation does not meet facility demand H-1.2.3: Inability to achieve synchronization	L-2, L-3, L-4
H-1.3: Incorrect sequencing of operations H-1.3.1: Out-of-Sync Re-closure H-1.3.2: Operation without permissive function H-1.3.3: Out-of-order switching operation	L-1, L-2, L-3, L-4
H.2: Inadequate monitoring and control of supply tie-line	
H-2.1: Plant is not isolated from supply tie-line during system fault	L-1, L-2, L-3, L-4
H-2.2: Plant is decoupled from tie-line when no fault exists	L-3, L-4
H-3: Inadequate isolation of faulted equipment	
H-3.1: Faulted components are not isolated correctly or fast enough	L-1, L-2, L-3, L-4
H-3.2: Isolation of no-fault area	L-3, L-4
H-4: Inadequate voltage transformation	
H-4.1: Operation in overloaded condition	L-1, L-2, L-3, L-4
H-4.2: Operation in overexcited condition	L-1, L-2, L-3, L-4

Table 4 - List of System-Level Constraints

System-Level Constraints	Related Hazards
SC-1: On-site generation must be adequately controlled/regulated	**H-1**
SC-1.1: Equipment must not be operated beyond normal operational limits SC-1.1.1: Mechanical parameters (speed, ramp-rates (acceleration), temperature, pressure, vibration as applicable) must not exceed normal operational limits SC-1.1.2: Electrical parameters (current, voltage, frequency) must not exceed normal operational limits	H-1.1 H-1.1.1 H-1.1.2
SC-1.2: Active and reactive power must be adequately controlled SC-1.2.1: Power quality metrics must not be violated SC-1.2.2: On-site generation must meet facility demand (in islanded mode) SC-1.2.3: Frequency, voltage and phase angle must be controlled to enable synchronization with the grid SC-1.2.4: Correct control mode must be used during operation	H-1.2 H-1.2.1 H-1.2.2 H-1.2.3 H-1.2.4
SC-1.3: Correct operating sequence must be followed SC-1.3.1: Out-of-sync re-closure must not occur SC-1.3.2: Permissive functions must not be violated SC-1.3.3: Switching function must follow correct order	H-1.3 H-1.3.1 H-1.3.2 H-1.3.3
SC.2: Supply tie-line must be adequately monitored and controlled	**H-2**
SC-2.1: Plant must be isolated from the supply tie-line during system fault	H-2.1
SC-2.2: Plant must not be decoupled from tie-line when no fault exists (when operating in grid mode)	H-2.2
SC-3: Faulted equipment must be isolated adequately	**H-3**
SC-3.1: Faulted components must be isolated quickly and in the correct order	H-3.1
SC-3.2: No-fault area must not be isolated	H-3.2
SC-4: Voltage transformation must be adequately controlled	**H-4**
SC-4.1: System must not operate in overloaded condition	H-4.1
SC-4.2: System must not operate in overexcited condition	H-4.2

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. Figure 5 shows the high-level functional control structure for the electric generation and distribution system.

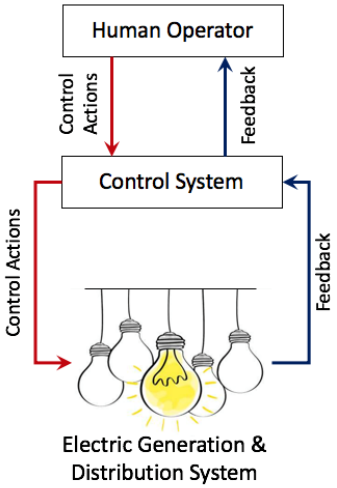

Figure 5 - High-Level Functional Control Structure

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. Figure 6 shows a refined version of the functional control structure. The figure shows a system of interacting controllers, primarily enforcing constraints on two controlled processes – *on-site generation* and *switching* 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). In its current configuration, RTAC is not part of the control structure at the plant; however, there are plans to install an RTAC system in the near future to enhance functionality by including 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.* [34], 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 5 and used as an input for the *extended functional control structure* presented in Figure 7. The figure demonstrates the dependency of the electric distribution system on natural gas, fuel oil, water as well as local electric utility distribution systems.

Table 5 - Identifying Interdependencies of the target system

Physical (Inputs/Outputs)
Natural gas at a min. pressure of 300 psig.
Fuel oil for backup stored on site and supplied via trucks
Lube oil for turbine, generator
Water for cleaning of compressor, emissions control
Purified Air
Electricity form the local grid
An infrastructure is physically dependent if there is a functional and structural linkage between the input and output of two assets: a commodity produced or modified by one infrastructure is required by another infrastructure for its operation
Cyber (Informational)
Various plant equipment communicates with external contractors for system monitoring via data link
Energy Management System (EMS) recommends throttle settings for turbine to optimize cost via internet
Plant information (PI) system records operational conditions in real-time and transmits to EMS as well as plant engineersover a DMZ router for business operations
An asset has a cyber dependency if its operation depends on information transmitted via electronic or informational links
Geographical (location)
Proximity of turbine to boilers & chillers
Fuel storage - gas/fuel oil
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
Logical
No logical interdependencies can be identified at this time
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

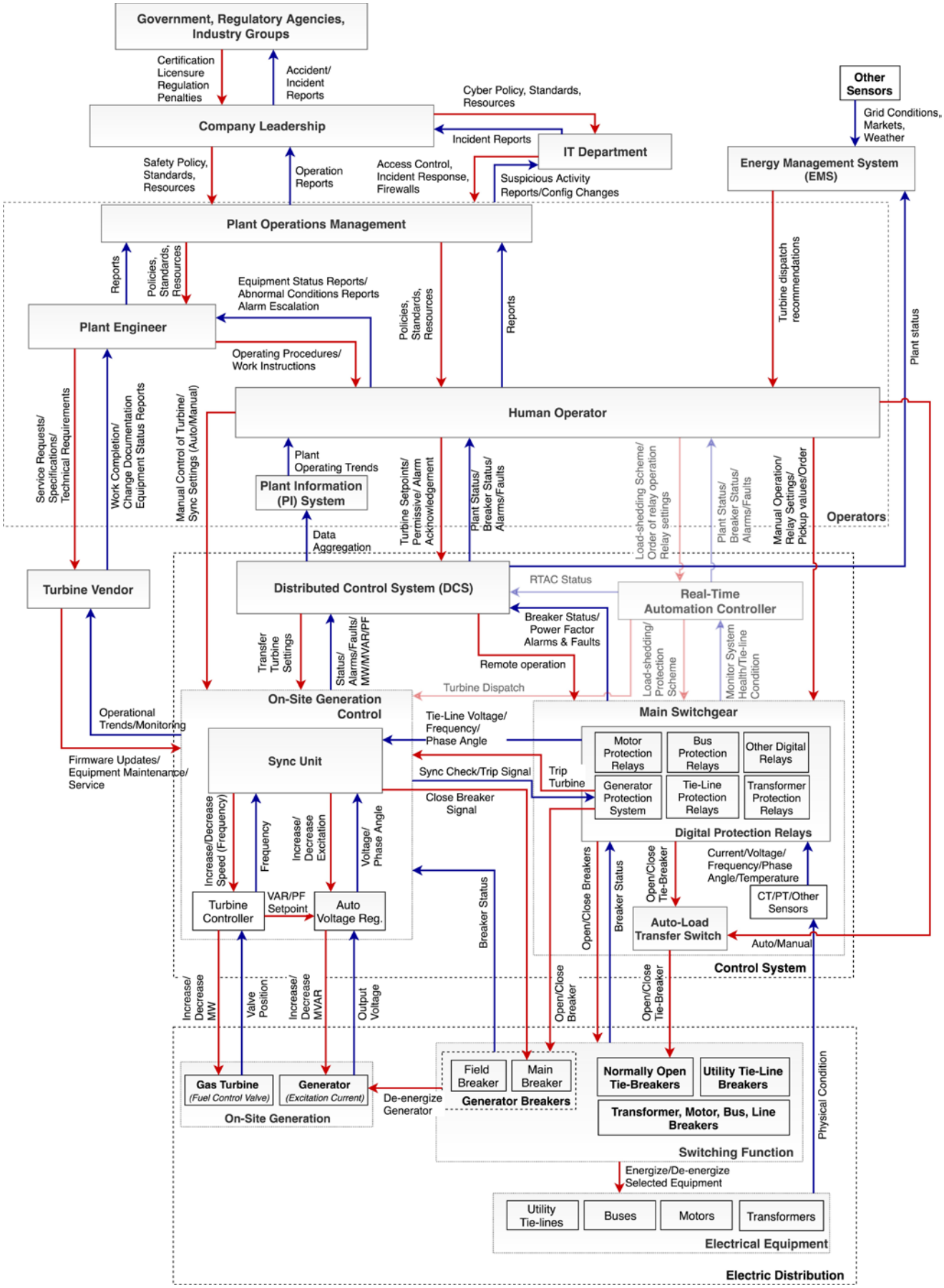

Figure 6 - Detailed Hierarchical Functional Control Structure

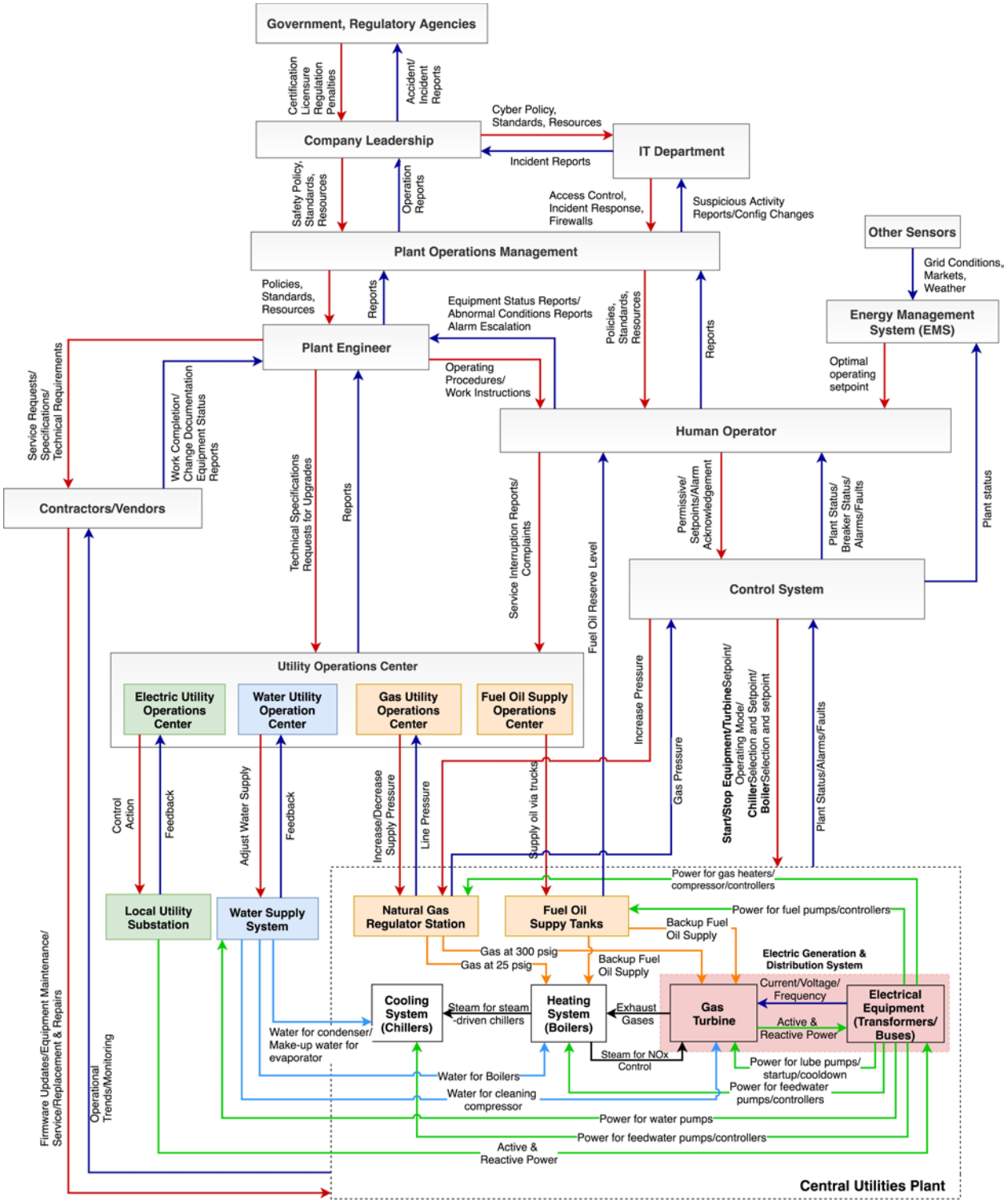

Figure 7 - Modeling System Interdependencies (physical and informational interdependencies) with Internal and External Systems

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 6.

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.

Table 6 - Partial List of Controllers, Safety Responsibilities & Control Actions

Controller	Function Performed	Safety Responsibilities	Control Actions
Operator	Perform day-to-day tasks to run equipment including the turbine, boilers and chillers in response to real-time demand variations from the MIT Campus	-Dispatch turbine, chillers, boilers to meet campus load -Monitor system operation for abnormalities -Emergency Shutdown equipment -Respond to alarms and faults and take corrective actions -Provide permissive functions and command overrides	-Select equipment and manually start/stop turbine and other equipment -Perform synchronization manually, by adjusting turbo-generator speed and excitation -Manually open/close breakers -Shutdown process during emergencies
Plant Engineer	The plant engineer is the technical lead for plant operations	-Ensure the operators have correct procedures -Ensure safety hazards are identified and mitigated -Verify equipment is functioning properly during operation -Ensure procedural compliance and training	-Approve operating procedures -Provide technical specifications and requirements to contractors/ vendors -Approve equipment change/modification requests -Troubleshooting
Sync Unit	Adjust speed and voltage of the turbine to enable synchronization with the grid	Ensure the frequency and phase angle of the generated power match that of the rest of the network	-Close gen. breaker when in-sync -Adjust turbine speed (frequency) -Adjust excitation (terminal voltage)
Turbine Controller	Maintain desired fuel flow to the turbine to match campus demand or operator provided set-points	-Control turbine active power output by adjusting fuel control valve to meet power demand -Enable Turbine Startup/Shutdown, synchronization including sequencing of all auxiliary equipment -Speed/load control, acceleration control, temperature control -Protect against overspeed, overtemperature, excessive vibration, loss of flame, loss of lube oil and other hazardous conditions	-Open fuel control valve (increase power output) -Close fuel control valve (decrease power output)
Automatic Voltage Regulator	Automatically maintain generator output terminal voltage at a set value under varying load and operating conditions; absorb or deliver reactive power for power factor (PF) control	-Control generator terminal voltage through adjustment of field current (excitation) or effectively, its reactive power output - Maintain system voltage when operating in islanded mode	-Increase excitation (rotor field current) -Decrease excitation (rotor field current)
Main Switchgear	Consists of a combination of electrical disconnect switches, fuses, circuit breakers and protective relays to control, protect and isolate electrical equipment; obtain metering and load information	-Protect equipment, service transformers and cabling against under/overcurrent, under/overvoltage and other unsafe conditions	
Digital Protection Relays	Microprocessor-based relay that analyses power system voltages, currents or other process quantities for the purpose of detection of faults in an electric power system; open/close remote bus tie breakers without operator intervention	-Primary responsibility is to immediately remove any individual component of a power system when it suffers a fault that might result in damage to property or unsafe conditions i.e. isolate faulty circuit from healthy circuit	-Trip breaker
Energy Management System	Combine real-time market and grid conditions along with predictive analytics to recommend operating points that maximize efficiency	-Provide recommendations within the generator capability limits	-Provide turbine dispatch recommendations

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 *Automatic Voltage Regulator* (AVR), is presented in Table 7.

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 [35]. Several UCAs for the AVR are listed in Table 8. 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.

Table 7 - 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

From Table 8 it is evident that the AVR performs a critical function in the stabilization of voltage and maintenance of power quality (or reactive power) 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"* [36][37][38]. The UCAs summarized in Table 8 are discussed next:

Table 8 - List of Unsafe Control Actions for the AVR

Action By	Control Action	Not Providing Causes Hazard	Providing Causes Hazard	Too soon, Too late, Out of order	Stopped too soon, Applied too long
AVR	**Increase/** Decrease Excitation	**UCA-AVR-1:** AVR does not provide excitation (loss of excitation) when coupled with the grid and operating at base load, causing the generator to operate as an induction generator, leading to overheating of rotor, insulation damage etc. --> [H-1.1.2, H-1.2.1]	**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-1.1.2]	**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, weakening the rotor magnetic field such that the generator 'slips a pole' causing catastrophic failure of the prime mover coupling --> [H-1.3; H-1.1.2]
		UCA-AVR-2: AVR does not increase excitation to match tie-line voltage preventing synchronization with the grid --> [H-1.2.3]	**UCA-AVR-6:** AVR increases excitation when generator terminal voltage is above setpoint (overvoltage) --> [H-1.1.2]		**UCA-AVR-10:** AVR increases excitation (too much) violating generator capability curve limits after synchronization with the grid, leading to rotor overheating --> [H-1.1]
		UCA-AVR-3: AVR does not increase excitation to achieve power factor or reactive power set-points, when synched with the grid, leading to utility financial penalties --> [H-1.2.1]	**UCA-AVR-7:** AVR increases excitation when turbine is tripped leading to high V/Hz (overfluxing) --> [H-1.1.2]		
		UCA-AVR-4: AVR does not increase excitation when system voltage falls below lower voltage limit triggering system voltage collapse --> [H-1.1.2]	**UCA-AVR-11:** AVR increases excitation when a large inductive load is removed causing overvoltage/overfluxing --> [H-1.1.2]		
		UCA-AVR-12: AVR does not increase excitation to required set-point when operating in islanded mode, causing undervoltage --> [H-1.1.2]			

a) Overexcitation

Overexcitation occurs whenever the ratio of the voltage to frequency (V/Hz) applied to the terminals of the generator exceed design limits [39], causing high flux density levels (i.e. overfluxing) in the generator core. *Mozina* [40] 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 [40]. 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 [41]. The overexcitation conditions can be caused by overvoltage, underfrequency or a combination of both [40] (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 [42] (UCA-AVR-11).

b) 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 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).

c) Overvoltage

Overvoltage occurs when the levels of electric field stress exceed the insulation capability of the generator stator windings [39]. 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 [42]. It can also stress insulation in other connected components such as transformers, bushings and surge arrestors.

d) 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 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 [40], [41] (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 [40], [41] (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 [43] (UCA-AVR-4).

e) 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 the motors [44] (UCA-AVR-12).

The systematic approach described above to identify UCAs for the AVR can be repeated for each of the other controllers modeled in the functional control structure as demonstrated by *Khan* [33].

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 Figure 8):

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 AVR from Figure 6 and superimpose it with guidewords from *Schmittner et al.* [32] signifying sample attack scenarios; the simplified control structure is presented in Figure 9. Starting with the AVR's *process model*, several causal factors are hypothesized which would cause the controller to issue an unsafe command. For instance, the controller could issue an unsafe command because it is fed *manufactured* 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 the same logic, each of the *sample attack guidewords* around the control loop are carefully contemplated as potential causal factors. New *constraints* and *functional requirements* are then defined to prevent the issuance of the UCA as well as to mitigate the effects of the attack.

Table 9 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 9) is provided in the next section. Meanwhile, we note that the last scenario in Table 9 (*Scenario SNR-AVR-08-06*) is unique because it identifies an *unsafe control input* originating from a *higher-level* controller i.e. the human operator. As before, two type of scenarios could cause the issuance of the unsafe control input; either the human operator provided an unsafe control input (i.e. pushing the wrong button) or took corrective action which was ineffective.

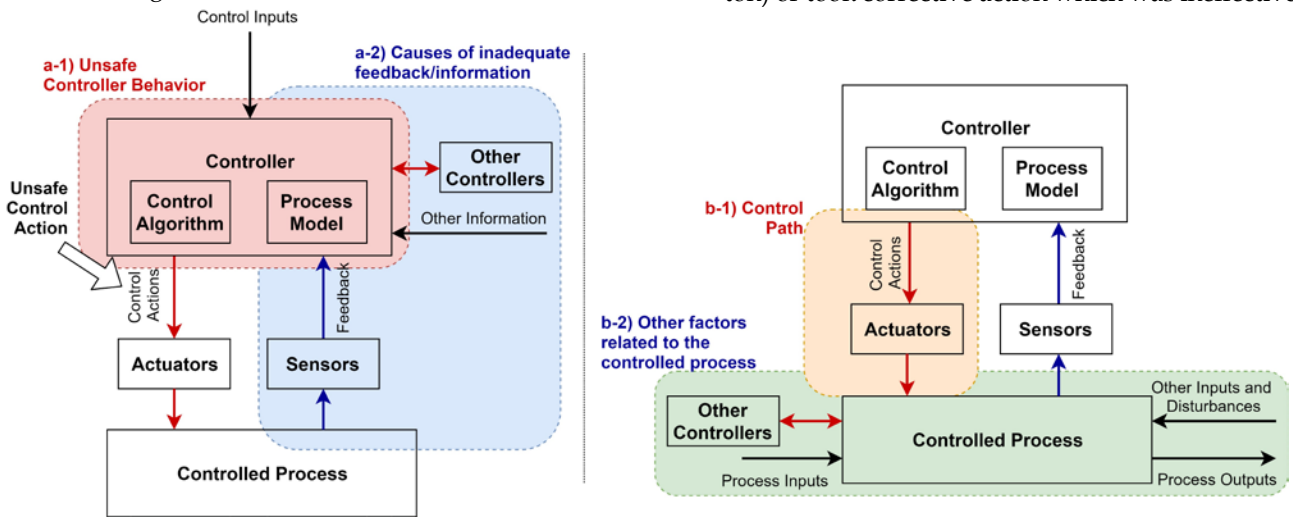

Figure 8 - Factors that can result in a) unsafe control actions b) safe control actions not or improperly executed

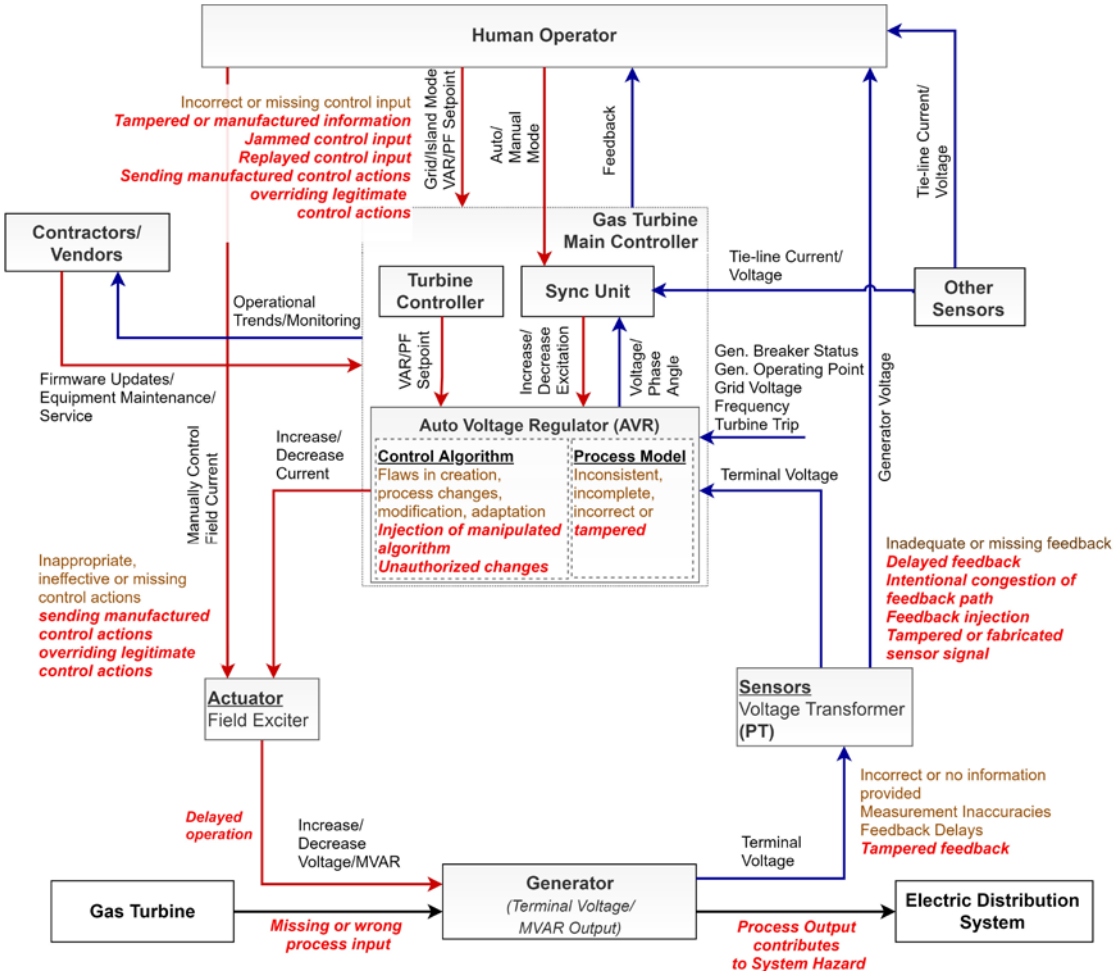

Figure 9 - Refined Control-Loop for the AVR

Table 9 - List of Scenarios for UCA-AVR-08

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] – (AVR Control Loop)

	Scenarios	Associated Causal Factors	Safety/Security Constraints
1	**No feedback or incorrect feedback** AVR does not know the actual state of the generator's terminal voltage; it continues to increase field excitation to achieve AFNL condition	1. Feedback path from the sensors (PT) is maliciously congested; AVR assumes previous state: a) delayed feedback b) no feedback received 2. Fabricated feedback signal is injected from the sensors (PT) to the AVR	1. AVR excitation signal must be interlocked with out-of-band feedback signal, such as turbine speed to prevent excitation beyond AFNL condition 2. AVR must alarm the operator, if at any time feedback is not received 3. Overflux Relay ANSI 24 must be installed with **separate** PT
2	**Malformed process model as a result of wrong info** AVR believes synchronization has already occurred and increases reference excitation signal to attain VAR/PF setpoint	1. Malicious feedback injection about status of generator breaker - AVR believes it to be closed when it is not	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	**Inadequate process model** While attempting to resynchronize with the grid, after a generator trip (but not a turbine trip), AVR increases excitation signal beyond AFNL	1. Breaker status does not get updated due to feedback drop or congestion of feedback path – AVR's process model is not updated to reflect loss of synchronization	1. AVR must have physical interlock with generator breaker - *Same as SC-AVR-08-02-01*
4	**Malformed control algorithm** AVR receives correct feedback about generator's terminal voltage but increases voltage to the wrong voltage setpoint due to unauthorized manipulation of AVR control algorithm	1. Contractor inadvertently uploads malicious firmware during routine maintenance which changes proportional-derivative-integral (PID) gain settings for voltage control	1. Unauthorized access to AVR must be prevented
5	**Control path compromised** AVR provides the correct signal but the field exciter (actuator) does not respond to the signal	1. Excitation mode is maliciously changed from *auto* to *manual* – unbeknownst to the operator, the control path is compromised and incorrect control signals are provided to the exciter by a malicious actor, masked as originating from the operator	1. AVR must have a physical switch to convert between *auto* and *manual* modes
6	**Incorrect control input from higher-level controller** AVR is provided the wrong voltage setpoint by the *Sync Unit* or the *operator*	1. Legitimate commands from the sync unit are jammed; manufactured commands from the sync unit/operator are sent to AVR 2. Incorrect setpoints are provided by higher-level controllers such as the operator or the sync unit (visit scenario UCA-OP-1)	*Same as SC-AVR-08-02-02 and SC-AVR-08-04-01*

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.

Figure 10 abstracts out some of the details in the control structure and presents a high-level view of the *Human Operator control loop*. As before, a number of causal factors are identified by evaluating the control loop and reasoning about the circumstances that would cause the transmission of an *unsafe command* to the AVR by this control loop. The loss scenarios and associated causal factors are summarized in Table 10. By following a similar approach, we can move around the control structure and evaluate each of the other controllers in order to generate a complete set of causal factors that lead to system-level losses

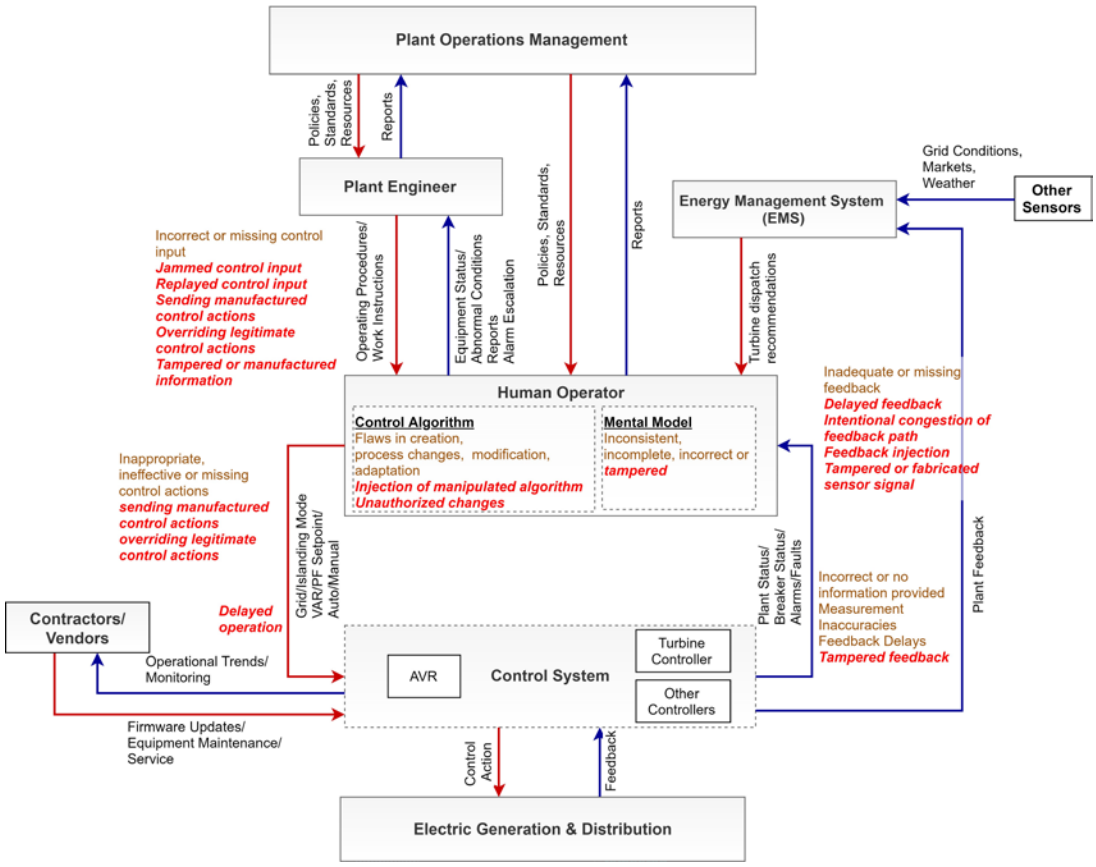

Figure 10 - Refined Control-Loop for the Operator

Table 10 - List of Scenarios for Operator

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] – (Operator Control Loop)			
	Scenarios	**Associated Causal Factors**	**Safety/Security Constraints**
6.1	**No feedback or incorrect feedback** Operator does not know the actual state of the generator's terminal voltage and believes the AVR has malfunctioned; overrides correct AVR signal (by forcing flags/overriding permissive functions) to manually increase excitation beyond AFNL setpoint	1. Feedback path from the sensors (PT) is maliciously congested or fabricated such that it is delayed or scattered; Operator is convinced AVR has malfunctioned, provides permissive functions to override AVR's legitimate signals which is then exploited by the attacker	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
6.2	**Incorrect control input from higher-level controller** 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	1. Operator is given the wrong procedure by the Plant Engineer	1. Standard Operating procedures must be secured 2. Plant Engineer must ensure operator has the correct procedure 3. Same as *SC-AVR-08-6.3-02* – hardcopy of procedure
6.3	**Malformed 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	1. AFNL set-points are maliciously altered in the operator's procedure because the procedure was stored in the unsecured part of the network 2. Operator is not adequately trained 3. Operator cannot access procedure because it is stored on the corporate network which is under a DOS attack – assumes an incorrect value	1. Operator must have hardcopy of quality controlled- procedure in the control room 2. Management must ensure operator is provided adequate training
6.4	**Control path hijacked** Operator knows the actual state of the generator's terminal voltage yet an incorrect control signal is provided to the field exciter	1. Operator is not adequately trained to translate an error code, seeks help on a public forum and inadvertently provides information about AVR model type and remote login credentials since a policy about sharing plant-specific information does not exist 2. Excitation mode is maliciously changed from auto to manual via the DCS; unbeknownst to the operator, the control path is compromised and incorrect control signals are provided to the exciter by a malicious actor 3. Malicious actor prevents lowering of excitation by taking control of the HMI and blocking operator access	1. Sharing of plant-specific information must be regularized via policy – this includes not just employees but also contractors and vendors 2. Unauthorized access to the DCS/Turbine Controller/AVR must be prevented 3. Out-of-band control and feedback must be provided to the operator - Same as SC-AVR-08-6.1-01 4. AVR must have a physical switch to convert between auto and manual modes

6 Discussion

The causal factors identified in Step 4 above (presented in Table 9 and Table 10) are indicative of different types of vulnerabilities. These include vulnerabilities that are local to the control loop being analyzed (missing component or 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.

For instance, the analysis revealed that in the event that the AVR malfunctions and causes the generator to overflux, the 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* [45] 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 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 9) 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. However, Scenarios SC-AVR-08-02 and -03 are interesting because they involve feedback about a component state that is not part of the AVR control loop – i.e. informational interdependency external to the control loop.

These flaws were discovered because the generator breaker status was identified as a variable that could affect the state of the AVR – making it a part of the process model of the AVR in Step 3. Scenario SC-AVR-08-03 is further unique because it involves a dynamic change in state that is exploited by a malicious actor. Both Scenarios SC-AVR-08-02 and -03 are examples of loss scenarios resulting from the interaction between components (the generator breaker and the AVR).

Scenario SC-AVR-08-04 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. Scenario SC-AVR-08-05 hypothesizes how legitimate control actions from the AVR can be hijacked by an attacker if the excitation mode selector has the ability to be remotely controlled (instead of through a physical selector switch).

Finally, Scenario SC-AVR-08-06 shows how unsafe conditions can emerge as a result of control inputs from hierarchical controllers. This scenario is refined into Scenarios SC-AVR-08-6.1 through -6.4 which each list causal factors where the operator either does not take corrective action or takes the wrong action (because of bad information) or takes the correct action which is not implemented successfully.

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-AVR-08-6.1 (Table 10) 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 provide permissive functions for manual override which is then exploited by the attacker. One important point to note here is that this scenario is possible because of poor cybersecurity culture, lack of cyber risk awareness etc., that may emerge as a result of management 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-AVR-08-6.2 is the result of the operator receiving an incorrect input (i.e. operating procedure) from a higher-level controller i.e. the plant engineer. Likewise, scenario SC-AVR-08-6.3 describes how the operator's control algorithm may be compromised by altering voltage set-points in the operating procedure – an indirect effect of not having access to physical copies of operating procedures in the control room. Scenario SC-AVR-08-6.4 describes how the lack of training and absence of policy for sharing plant specific information can enable an attacker to gain a foothold in the network that can be exploited later.

The bottom part of Figure 9 highlights sample attack scenarios that emerge as a result of interdependencies of the controlled process with other components. Similarly, the last row in Table 9 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.

Note that despite the limited nature of the analysis, different types of vulnerabilities have been uncovered. Figure 11 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 Figure 11 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 for 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.

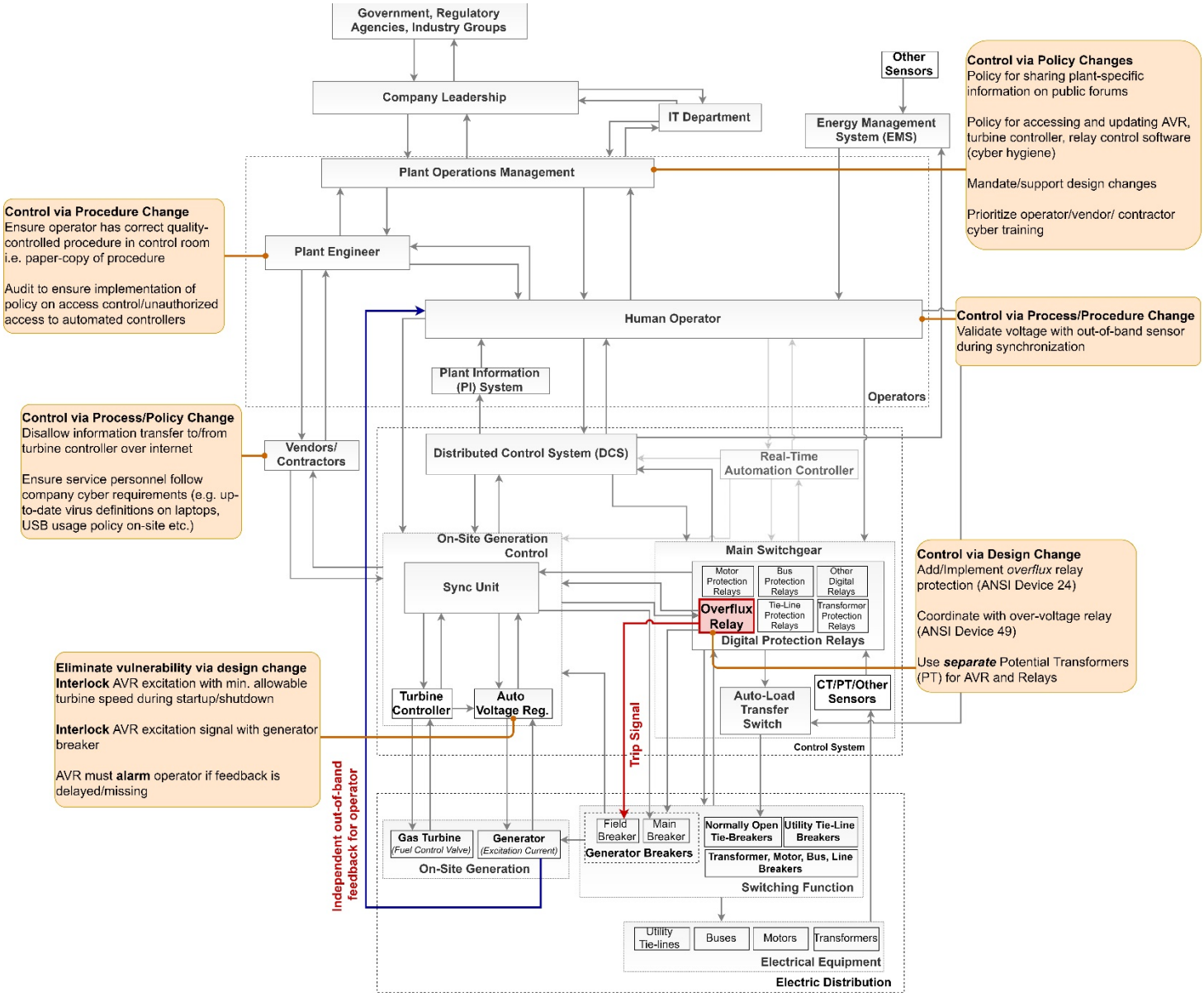

Figure 11 - A subset of proposed requirements/constraints to eliminate and/or mitigate vulnerabilities

7 Conclusion

The objective of this study to demonstrate the *Cybersafety* method in systematically and robustly uncovering cyber vulnerabilities and mitigation strategies in an industrial control system; 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 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.

Acknowledgment

This material is based, in part, upon research supported by the Department of Energy under Award Number DE-OE0000780, a seed grant from the MIT Energy Initiative (MITei), and funds from 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, n.d.

[2] M.G. Angle, S. Madnick, J.L. Kirtley, Identifying and Anticipating Cyber Attacks that could cause Physical Damage to Industrial Control Systems, 2017. http://web.mit.edu/smadnick/www/wp/2017-14.pdf (accessed May 4, 2019).

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

[4] GE Grid Solutions, STV Overexcitation Relay, (n.d.). https://www.gegridsolutions.com/multilin/catalog/stv.htm (accessed December 19, 2019).

[5] General Electric LM2500 | PowerWeb, (n.d.). http://www.fi-powerweb.com/Engine/Industrial/GE-LM2500.html (accessed December 19, 2019).

[6] A.A. Cárdenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, S. Sastry, Challenges for Securing Cyber Physical Systems, n.d. https://ptolemy.berkeley.edu/projects/chess/pubs/601/cps-security-challenges.pdf (accessed May 1, 2019).

[7] C. Johnson, Why We Cannot (Yet) Ensure the Cyber-Security of Safety-Critical Systems, in: 24th Safety-Critical Syst. Symp., Brighton, UK, 2016: pp. 171–182. http://www.dcs.gla.ac.uk/~johnson,Johnson@dcs.gla.ac.uk (accessed May 1, 2019).

[8] C. Schmittner, T. Gruber, P. Puschner, E. Schoitsch, Security Application of Failure Mode and Effect Analysis (FMEA), in: Springer, Cham, 2014: pp. 310–325. https://doi.org/10.1007/978-3-319-10506-2_21.

[9] H.A. Duckworth, R.A. Moore, Social responsibility: Failure mode effects and analysis, CRC Press, 2010.

[10] M. Steiner, P. Liggesmeyer, Combination of Safety and Security Analysis - Finding Security Problems That Threaten The Safety of a System, in: 32nd Int. Conf. Comput. Safety, Reliab. Secur., France, 2013. http://conference.hitb.org/hitbsecconf2013ams/hugo-teso/ (accessed May 1, 2019).

[11] H.A. Watson, Launch control safety study, Bell Labs. (1961).

[12] A. Altawairqi, M. Maarek, Attack Modeling for System Security Analysis, in: Springer, Cham, 2017: pp. 81–86. https://doi.org/10.1007/978-3-319-66284-8_8.

[13] C.A. Ericson, Hazard Analysis Techniques for System Safety, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005. https://doi.org/10.1002/0471739421.

[14] L. Dawson, A. Muna, T. Wheeler, P. Turner, G. Wyss, 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, n.d. https://www.osti.gov/servlets/purl/1252915 (accessed May 1, 2019).

[15] X. Xu, M.L. Ulrey, J.A. Brown, J. Mast, M.B. Lapis, Safety Sufficiency for NextGen Assessment of Selected Existing Safety Methods, Tools, Processes, and Regulations, 2013. http://www.sti.nasa.gov (accessed May 2, 2019).

[16] N.G. Leveson, Is Estimating Probabilities the Right Goal for System Safety?, n.d. http://sunnyday.mit.edu/papers/Making-Safety-Decisions.pdf (accessed October 7, 2019).

[17] J. Dunjó, V. Fthenakis, J.A. Vílchez, J. Arnaldos, Hazard and operability (HAZOP) analysis. A literature review, J. Hazard. Mater. 173 (2010) 19–32. https://doi.org/10.1016/j.jhazmat.2009.08.076.

[18] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, S. Sezer, STPA-SafeSec: Safety and security analysis for cyber-physical systems, J. Inf. Secur. Appl. 34 (2017) 183–196. https://doi.org/10.1016/J.JISA.2016.05.008.

[19] N.G. Leveson, J.P. Thomas, STPA Handbook, 2018. http://psas.scripts.mit.edu/home/ (accessed April 28, 2019).

[20] N. Leveson, Engineering a safer world : systems thinking applied to safety, The MIT Press, 2012.

[21] J.P. Thomas IV, Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis, (2013).

[22] R. Sotomayor Martínez, System Theoretic Process Analysis of Electric Power Steering for Automotive Applications, 2015. http://sunnyday.mit.edu/STAMP/Sotomayor-Thesis.pdf (accessed May 2, 2019).

[23] N.A. Peper, Systems Thinking Applied to Automation and Workplace Safety Signature of Author, MIT, 2007. http://sunnyday.mit.edu/peper-thesis.pdf (accessed May 2, 2019).

[24] J. Thomas, F. Luiz De Lemos, N. Leveson, Evaluating the Safety of Digital Instrumentation and Control Systems in Nuclear Power Plants, 2012. http://sunnyday.mit.edu/papers/MIT-Research-Report-NRC-7-28.pdf (accessed May 2, 2019).

[25] P.D. Stukus, Systems-Theoretic Accident Model and Processes (STAMP) Applied to a U.S. Coast Guard Buoy Tender Integrated Control System, 2017. http://sunnyday.mit.edu/stukus-thesis.pdf (accessed May 1, 2019).

[26] T. Pawlicki, A. Samost, D.W. Brown, R.P. Manger, G.-Y. Kim, N.G. Leveson, Application of systems and control theory-based hazard analysis to radiation oncology, Med. Phys. 43 (2016) 1514–1530. https://doi.org/10.1118/1.4942384.

[27] J.R. Laracy, A Systems-Theoretic Security Model For Large Scale, Complex Systems Applied To The Us Air Transportation System, 2007. https://dspace.mit.edu/bitstream/handle/1721.1/39256/173417210-MIT.pdf?sequence=2&isAllowed=y (accessed May 2, 2019).

[28] J.R. Laracy, Applying STAMP to Critical Infrastructure Protection, 2007. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.7643&rep=rep1&type=pdf (accessed May 2, 2019).

[29] H.M. Salim, Cyber safety : a systems thinking and systems theory approach to managing cyber security risks, (2014). https://dspace.mit.edu/handle/1721.1/90804 (accessed May 21, 2019).

[30] A. Nourian, S. Madnick, A Systems Theoretic Approach to the Security Threats in Cyber Physical Systems Applied to Stuxnet, IEEE Trans. Dependable Secur. Comput. 15 (2018) 2–13. https://doi.org/10.1109/TDSC.2015.2509994.

[31] W. Young, N.G. Leveson, An integrated approach to safety and security based on systems theory, Commun. ACM. 57 (2014) 31–35. https://doi.org/10.1145/2556938.

[32] C. Schmittner, Z. Ma, P. Puschner, Limitation and Improvement of STPA-Sec for Safety and Security Co-analysis, in: Springer, Cham, 2016: pp. 195–209. https://doi.org/10.1007/978-3-319-45480-1_16.

[33] S. Khan, Using a System-Theoretic Approach to Identify Cyber-Vulnerabilities and Mitigations in Industrial Control Systems Submitted to the System Design and Management Program in Partial Fulfillment of the Requirements for the Degree of, Massachusetts Institute of Technology, 2019. http://dspace.mit.edu/handle/1721.1/7582 (accessed October 15, 2019).

[34] S.M. Rinaldi, J.P. Peerenboom, T.K. Kelly, Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies, n.d. https://pdfs.semanticscholar.org/b1b7/d1e0bb39badc3592373427840a4039d9717d.pdf (accessed April 27, 2019).

[35] J. Thomas, Extending and automating a Systems-Theoretic hazard analysis for requirements generation and analysis., (2012). https://doi.org/10.2172/1044959.

[36] Guard against over-fluxing: Ensure proper generator protection, maintenance, (2015). http://www.ccj-online.com/guard-against-over-fluxing-ensure-proper-generator-protection-maintenance/ (accessed April 23, 2019).

[37] VOITH, Thyricon™ Excitation System, 2016. https://voith.com/corp-en/VH_Product_Brochure_Automation_Thyricon_14_vvk_t3387_en.pdf (accessed April 23, 2019).

[38] GE Automation & Controls, EX2100e Excitation Control 100 mm, 77 mm, 53 mm, and 42 mm Thyristor Systems Product Description, 2013. http://www.geautomation.com/system/files/files/gea-s1302.pdf (accessed April 23, 2019).

[39] IEEE IPES Generator Protection Task Force, IEEE Tutorial on the Protection of Synchronous Generators, 1995. http://resourcecenter.ieee-pes.org/pes/product/technical-publications/PESTP102 (accessed April 24, 2019).

[40] C.J. Mozina, Power plant "horror stories," in: Proc. Inaug. IEEE PES 2005 Conf. Expo. Africa, IEEE, n.d.: pp. 462–465. https://doi.org/10.1109/PESAFR.2005.1611865.

[41] G. Klempner, I. Kerszenbaum, Operation and Maintenance of Large Turbo-Generators, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2004. https://doi.org/10.1002/0471683388.
[42] R.C. Scharlach, J. Young, Lessons Learned From Generator Event Reports, 2010. https://cdn.selinc.com/assets/Literature/Publications/Technical Papers/6387_LessonsLearnedGeneratorRS-JY_20100303_Web.pdf?v=20151125-164158 (accessed April 24, 2019).
[43] W. Hartmann, Generator Protection Overview, n.d. https://www.eiseverywhere.com/file_uploads/8b5452d7f9376912edcba156cd1a5112_WSU_GENPROTOVERVIEW_180305.pdf (accessed April 25, 2019).
[44] UST Power, AVR Guide: Voltage Too High, Too Low | UST, (n.d.). https://ustpower.com/comparing-automatic-voltage-regulation-technologies/need/avr-guide-voltage-high-low/ (accessed April 24, 2019).
[45] R.C. Scharlach, J. Young, Lessons Learned From Generator Event Reports, 2010. https://cdn.selinc.com/assets/Literature/Publications/Technical Papers/6387_LessonsLearnedGeneratorRS-JY_20100303_Web.pdf?v=20151125-164158 (accessed May 6, 2019).

Shaharyar Khan is a research scientist at the MIT Sloan School of Management. He received his S.M in Engineering and Management from the Massachusetts Institute of Technology. He received his BASc Hons. (2010) degree in Mechanical Engineering from the University of Waterloo. He has worked as a seismic/structural design engineer for BWX Technologies, designing and analyzing critical components for nuclear power plants. He has also worked as a Lead Project Engineer at Bruce Power Nuclear Generating Station, deploying tools for reactor inspections and maintenance. He is a Registered Professional Engineer in Ontario, Canada.

Stuart Madnick is the John Norris Maguire (1960) Professor of Information Technology and a Professor of Engineering Systems at the Massachusetts Institute of Technology. He has been an MIT faculty member since 1972. He has served as the head of MIT's Information Technologies Group in the MIT Sloan School of Management for more than twenty years. He is the Founding Director of Cybersecurity at MT Sloan (CAMS): The Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity. He is the author or co-author of over 350 books, articles, or reports including the classic textbooks, Operating Systems (1974) and Computer Security (1979). His current research interests include cybersecurity, information integration technologies, semantic web, software project management, internet applications, and the strategic use of information technology. Madnick 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 has served as a consultant to major corporations, the founder or co-founder of five high-tech firms, and currently operates a hotel in the 14th century Langley Castle in England. Madnick holds an SB in electrical engineering, an SM in management, and a PhD in computer science from MIT.