SAND2019-6969C

Addressing Cyber Hazards in Nuclear Power Plants with Systems Theoretic-Informed Fault Trees

Adam D. Williams & Andrew J. Clark

Sandia National Laboratories, Albuquerque, NM, USA, [adwilli; ajclark]@sandia.gov*

Civilian nuclear applications—including nuclear power plants (NPPs)—are trending towards modernizing plant control systems from analog to digital instrumentation and control (DI&C) systems. Though well established and mature, traditional probabilistic risk assessment (PRA) methods for NPP safety analyses struggle to adequately address vulnerabilities introduced by digital equipment and other cyber hazards. More specifically, the potential failures and/or undesired behaviors due to plant modernization will manifest from digital and passive systems—whose behaviors are not as aligned with core tenets of reliability theory—as older NPPs who relied on analog and active systems. Additionally, traditional risk assessment tools do not account for systems that perform their functions but still lead to inadequate behavior.

Recent research sponsored by the Electric Power Research Institute (EPRI) has aimed to rectify this struggle. This research has shown that the logical process for prioritizing the importance of component behavior within varying loss scenarios of fault tree analysis (FTA) can be combined with the top-down process for evaluating emergent behaviors of systems-theoretic process analysis (STPA). The results are so-called "STPA-informed fault trees" (SIFTs), which have emerged as a powerful analytical tool for evaluating cyber hazards in NPPs. By incorporating STPA-derived hazardous control actions into fault trees, the resulting SIFT cut sets can be categorized in terms of whether they contact only physical, only digital, or a combination of digital/physical components. This provides unique insights into developing and managing cyber protection strategies. This hybrid analytical approach has been codified into a process called Hazards and Consequences Analysis for Digital Systems (HAZCADS) which seeks to leverage the respective benefits of both FTA and STPA approaches.

After introducing the challenges of digital controllers and cyber hazards to desired NPP operations, this paper will briefly review the core tenets of both FTA and STPA. Next, both a detailed description and example of how SIFTs are used to evaluate cyber hazards in NPPs will be provided. Finally, this paper will summarize the overall HAZCADS methodology and offer insights for improving cyber security efforts for civilian nuclear applications.

INTRODUCTION

Civilian nuclear applications—including nuclear power plants (NPPs)—are trending towards modernizing plant control systems from analog to digital instrumentation and control (DI&C) systems.[1] The inclusion of digital instrumentation and control (DI&C) into nuclear power plants (NPP) presents new and unique challenges to traditional risk analysis approaches. DI&C may be accompanied by non-traditional failure modes, such as design errors, software flaws, and cyber-attack threats.[2] Due to the prevalence of DI&C in NPPs and the growing threat of cyber attacks, it is imperative to improve digital hazard analysis for NPPs.

* **SAND2019-XXXX C**. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Though well established and mature, traditional probabilistic risk assessment (PRA) (e.g., [3]) methods for NPP safety analyses struggle to adequately address vulnerabilities introduced by digital equipment and other cyber hazards. Potential failures and/or undesired behaviors due to plant modernization will manifest from digital and passive systems—whose behaviors are not as aligned with core tenets of reliability theory—as older NPPs who relied on analog and active systems. More specifically, traditional PRAs assess failure modes of process components (e.g., pump failure) and operator actions. DI&C failure modes, however, may also manifest as systematic failures—or, *non-traditional failure modes*—that result from complex interactions are not readily included into PRA models. Additionally, traditional risk assessment tools do not account for systems that perform their functions but still lead to inadequate or unexpected behavior.

LEVERAGING NEW INSIGHTS FOR DIGITAL HAZARDS ANALYSIS

Recent research sponsored by the Electric Power Research Institute (EPRI) has aimed to rectify this struggle. During this research, EPRI and Sandia National Laboratories (Sandia) developed an evaluation rubric to advance the use of hazards analysis methods to assess cyber vulnerabilities. [4][5][6] The elements of this evaluation criteria included:

- Determine a "holistic" characterization of the NPP;
- Prioritize risk for a "holistic" system characterization;
- Identify new failure modes unique to DI&C components;
- Describe new interactions enabled by DI&C design features;
- Illustrate new system effects from DI&C-related failure modes and interactions; and,
- Visualize the interrelationships between DI&C and non-DI&C system elements.

This evaluation rubric was applied to a suite of traditional and novel hazard analysis techniques in order to identify non-traditional failure modes that compromise the intended system control and provides a means for consequence analysis and risk prioritization for highly digital systems. One of the conclusions of this research was that no single methodology, in its current form, is befitting to address all the potential and emerging hazards and consequences for digital systems in NPPs. Furthermore, the EPRI reports suggested that instead of attempting to invent a new methodology for addressing digital systems, two of the established methodologies can be combined into a unified methodology.[4][5]

Two current hazard analysis techniques—System-Theoretic Process Analysis (STPA) and Fault Tree Analysis (FTA)—individually measured well against these criteria. Yet, their potential *combination* was further explored and demonstrated an even higher capability to meet the criteria necessary for evaluating risk-informed cyber security at NPPs. More specifically, EPRI's research indicated that the logical process for prioritizing the importance of component behavior within varying loss scenarios of fault tree analysis (FTA) can be combined with the top-down process for evaluating emergent behaviors of systems-theoretic process analysis (STPA).[4][5]

Fault Tree Analysis (FTA)

Fault trees are *bottom-up*, deductive logic models for complex systems. Development of a FTA model starts by defining the occurrence of a top event representing an undesirable outcome for a facility or process (e.g., release of chemicals to environment, failure to provide electrical power) or even simply a single system or set of several systems (e.g., loss of primary coolant system integrity). FTA is used to identify combinations of failure modes of structures, sub-systems and components that lead to failure of systems to perform their intended functions. FTA has been applied as a method to study NPP system design for over fifty years.[7] Fault trees can use Boolean equations to quantify the failure probability of a system or collection of systems—or, conversely, estimate their reliability—through the implementation of probabilities of the primary events. However, the utility of FTA is not exclusive to quantitative probabilistic results. Fault trees yield equally as useful qualitative insights regarding the design of a plant and its systems.

Systems-Theoretic Process Analysis (STPA)

Systems Theoretic Process Analysis (STPA) is a *top-down* hazard analysis method that is part of a relatively new set of system safety methods being developed at the Massachusetts Institute of Technology (MIT) [8]. More specifically, STPA describes how undesired outcomes (e.g., losses) can result from inadequate enforcement of constraints (e.g., control) on design, development, and operation of systems to achieve desired objectives. This logical perspective asserts that system losses result from flawed interactions between physical components, engineering activities, operational mission, organizational structures and social factors. The strengths of STPA include its ability to identify undesired outcomes that arise from dysfunctional component interactions and incorporate realistic descriptions of human influences—versus the simple component failures or assignment "human error" rates in traditional approaches.

SYSTEMS-THEORETIC INFORMED FAULT TREES (SIFTs)

Incorporating unsafe control actions into fault tree models leads to a fundamentally new model called "systems-theoretic informed fault trees," or SIFTs. SIFTs better incorporate both the direct and indirect roles of digital components in potential failure pathways. Here, SIFTs expand upon traditional fault trees by incorporating (1) the uniqueness and complexity of DI&C components and (2) newly identified causes of hazards ("failures" in traditional FTA terminology) including those from component interactions and that still result with no component failure occurring.

By incorporating STPA-derived hazardous control actions into fault trees, the resulting SIFT cut sets can be categorized in terms of whether they contact only physical, only digital, or a combination of digital/physical components. This provides unique insights into developing and managing cyber or digital hazard mitigation and management strategies.[6] More specifically, incorporating digital components into fault tree models allows the analytic power of Boolean algebra to describe the impact of such components on system-level behaviors.

Solving these fault tree models results in three specific categories of potential cut sets (i.e., combinations of events that result in a hazard). First, or *Type I*, are cut sets comprised solely of non-digital hardware component failures (and, therefore, those identified with traditional FTA). Second, or *Type II*, are cut sets comprised of combinations of unsafe control actions from digital components with non-digital hardware component failures. Third, or *Type III*, are cut sets comprised only of unsafe control actions from digital components. SIFTs better incorporate both the direct and indirect roles of digital components in potential failure pathways and expand upon traditional fault trees by incorporating: (1) the uniqueness and complexity of DI&C components; and (2) newly identified digital failure modes, including those from component interactions that still result with no component failure occurring.

Both Type II and III can identify where mitigation measures might need to be implemented. Additionally, SIFTs can identify digital I&C components that have no impact on the loss or top event. The implication of this finding is that limited resources can be focused on more hazardous digital components. Type 3 cut sets reveal potential digitally related faults or cyber exploitation vulnerabilities that result in system failure through the DI&C system. Furthermore, Type 3 cut sets will identify control actions from a single digital component. If hazardous control actions associated with specific digital assets occur only in Type 2 cut sets (i.e., combinations of digitally related basic events and non-digital basic events), then those digital assets represent new opportunities or mitigation measures for attackers of the digital systems.

SIFT EXAMPLE: LabVolt Bench-Scale System

For demonstration, Sandia's LabVolt System (LabVolt) bench-scale system was used to investigate the impact of cyber hazards on a physical system. The function of the LabVolt System is to circulate water through a heating, ventilation and air conditioning (HVAC) cooling tank.[1] As shown **Error! Reference source not found.**, the pump sends water to the HVAC cooling tank and the valve regulates the water level in the HVAC cooling tank by opening or closing. The fan ensures that the HVAC cooling tank does not exceed its operational temperature. There is also a supply tank that feeds the pump and fills when the valve operates. In addition, the LabVolt system also includes the digital architecture –including several digital components, controllers, and a human machine interface (HMI)—shown in Figure 2.

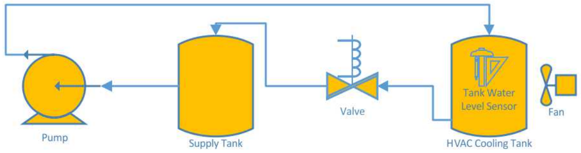

Figure 1. LabVolt lab bench-scale system P&ID.

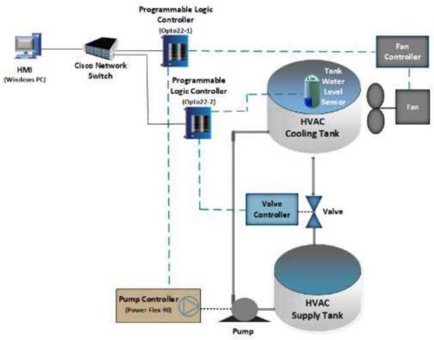

Figure 1. LabVolt digital network and process equipment.

According to the tenets of STPA, the physical elements of Figure 1 and digital elements of Figure 2 were translated into a hierarchical control structure for the LabVolt system. For clarity, Figure 3 illustrates a hybrid HCS for the LabVolt system—wherein the combination of controllers, control actions (CA) and feedbacks (FB) are overlaid on top of a more traditional process-based system representation. (NOTE: This is a useful intermediate step to help increase the utility and understandability of HCS to unpracticed users of STPA.) The control actions (CA) and feedbacks (FB) are labeled in Figure 3 to describe the interactions between physical, digital, and human components within the LabVolt system. (NOTE: Because of the emphasis on digital exploits in cyber security planning feedback has been separated into analog (e.g., electric current going through a wire) and digital (e.g., signal sent through an ethernet connection) feedback labeled in Figure 3 as FBA and FBD, respectively.)

[1] There is no actual process heat in the LabVolt system. The only actual process variable is the HVAC water level. Regardless, the system was designed as a heat removal system, thus, throughout this section are references to the cooling function of the system.

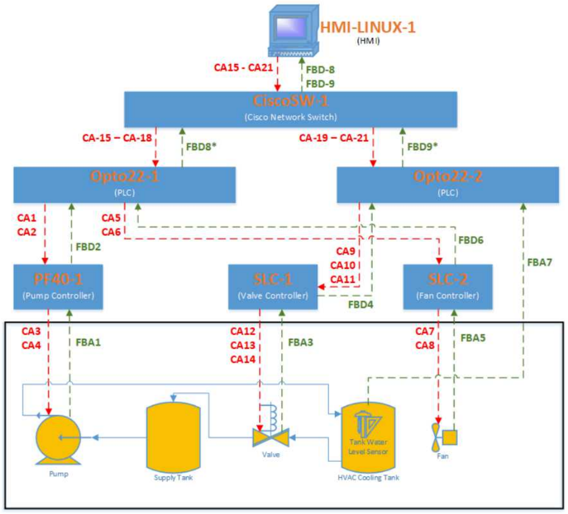

Figure 2. LabVolt hierarchical control structure with digital components overlaid with process equipment.

Using STPA, each version of inadequate control actions (or feedbacks) is described in terms of how that loss of control can lead to a specific hazardous state. For example, control actions typically considered adequate (or at least benign) can result in hazardous states when issued (i.e., the "provided, not needed" condition) or issued at an incorrect temporal or sequential point. SIFTs help evaluate these hazardous control actions to identify the combination of process component failures that lead to a top-level event occurring. SIFTs visualize the interconnection of STPA and FTA results to identify the gate(s) in the fault tree where hazardous control actions are inserted. As shown in Figure 4, for example, if the LabVolt pump fails to run (e.g., due to pump shaft failure), then one of two process component failures—the other being valve fails open—is satisfied.

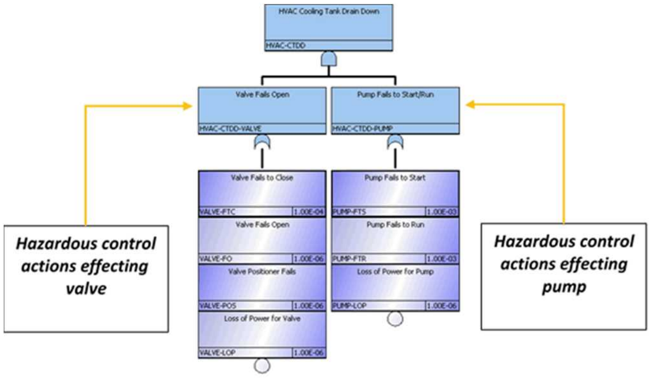

Figure 4. Illustration of where hazardous control actions will be added to a fault tree for HAZCADS of the LabVolt system.

More specifically, Figure 5 illustrates the construction of the SIFT[2] by adding all hazardous control actions identified as contributing to the hazard (or the fault tree top event) *HVAC Cooling Tank Drain Down* in Figure 4. Using a traditional fault tree for this hazard has seven basic events, whereas

[2] Due to the expansion of the fault tree, the SIFT in Figure 5 is difficult to read and for clarity a zoomed in portion of the lower gate Valve Fails Open is provided in Figure 6 to highlight specific features of the SIFT 6.

the SIFT has 30 events. The top event (*HVAC Cooling Tank Drain Down*), the valve and pump gates (*Valve Fails Open* and *Pump Fails to Start/Run*, respectively), along with the basic events remain unchanged.

New features of the SIFT result from adding new fault tree gates for each controller in the system, along with a send or do not send signal that ultimately leads to a valve being open to better model digitally related faults. In Figures 5 & 6, hazardous control actions are separated into "provided, not needed (PNN)" and "needed, not provided (NNP)" gates. Underneath each of these gates are the hazardous control action events that if performed, can create the hazard. Each of the hazardous control action events are represented by an "undeveloped event". Fault tree undeveloped events are, as the name implies, not fully resolved with respect to a defined failure mode. As such, the SIFT represents a more comprehensive description that includes both digital assets and a more accurate, contextualized model of system operations.

Although the cut sets are ultimately assessed qualitatively, traditional fault tree analysis can be used to readily identify the three types of cut sets. For example, one approach is to imitate probabilities for the hazardous control actions. If they are assigned a value of one in the fault tree software then post-processing of the results can categorize all cut sets with hazardous control actions quantitatively. This approach is used for the post-processing of cut sets for the LabVolt system. Solving the systems-theoretic informed fault tree in Figure 5 leads to a total of 176 SIFT cut sets. In addition, examples of Type 1, Type 2 and Type 3 SIFT cut sets are presented in Table 1, including illustrating how Type 2 and specifically Type 3 cut sets correspond to faults of digital components.

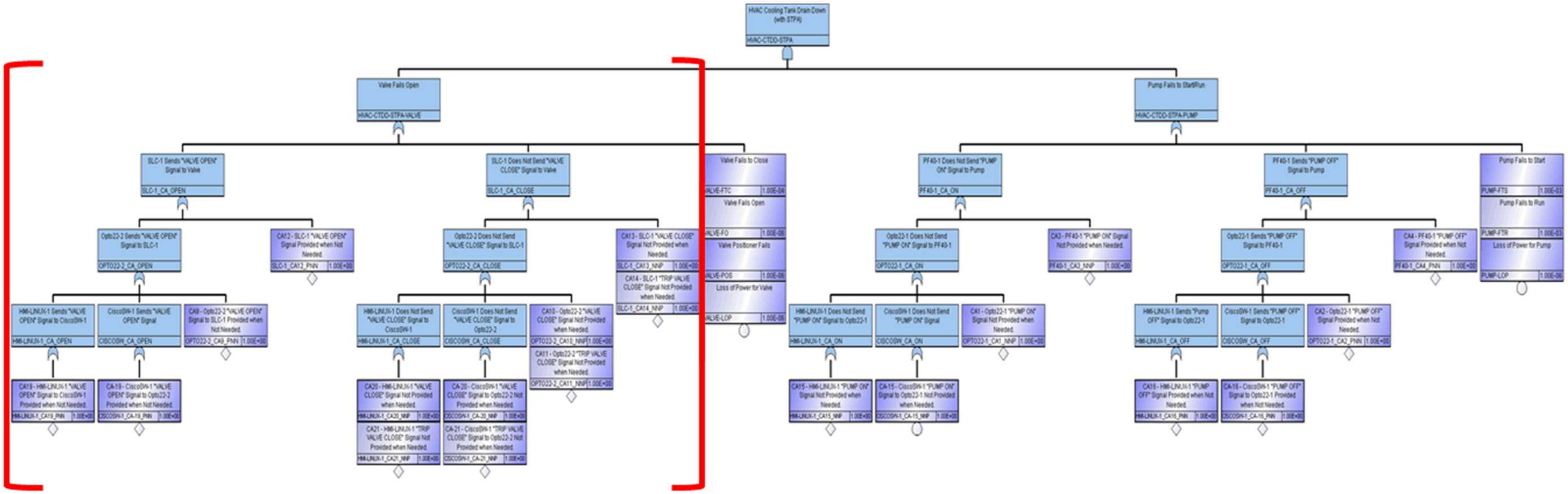
HVAC Cooling Tank Drain Down (with STPA)
HVAC-CTDD-STPA
Valve Fails Open
HVAC-CTDD-STPA-VALVE
SLC-1 Sends "VALVE OPEN" Signal to Valve
SLC-1_CA_OPEN
SLC-1 Does Not Send "VALVE CLOSE" Signal to Valve
SLC-1_CA_CLOSE
Opto22-2 Sends "VALVE OPEN" Signal to SLC-1
OPTO22-2_CA_OPEN
CA12 - SLC-1 "VALVE OPEN" Signal Provided when Not Needed.
SLC-1_CA12_PNN 1.00E+00
Opto22-2 Does Not Send "VALVE CLOSE" Signal to SLC-1
OPTO22-2_CA_CLOSE
CA13 - SLC-1 "VALVE CLOSE" Signal Not Provided when Needed.
SLC-1_CA13_NNP 1.00E+00
CA14 - SLC-1 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
SLC-1_CA14_NNP 1.00E+00
HMI-LINUX-1 Sends "VALVE OPEN" Signal to CiscoSW-1
HMI-LINUX-1_CA_OPEN
CiscoSW-1 Sends "VALVE OPEN" Signal
CISCOSW_CA_OPEN
CA9 - Opto22-2 "VALVE OPEN" Signal to SLC-1 Provided when Not Needed.
OPTO22-2_CA9_PNN 1.00E+00
HMI-LINUX-1 Does Not Send "VALVE CLOSE" Signal to CiscoSW-1
HMI-LINUX-1_CA_CLOSE
CiscoSW-1 Does Not Send "VALVE CLOSE" Signal to Opto22-2
CISCOSW_CA_CLOSE
CA10 - Opto22-2 "VALVE CLOSE" Signal Not Provided when Needed.
OPTO22-2_CA10_NNP 1.00E+00
CA11 - Opto22-2 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
OPTO22-2_CA11_NNP 1.00E+00
CA19 - HMI-LINUX-1 "VALVE OPEN" Signal to CiscoSW-1 Provided when Not Needed.
HMI-LINUX-1_CA19_PNN 1.00E+00
CA-19 - CiscoSW-1 "VALVE OPEN" Signal to Opto22-2 Provided when Not Needed.
CISCOSW-1_CA-19_PNN 1.00E+00
CA20 - HMI-LINUX-1 "VALVE CLOSE" Signal Not Provided when Needed.
HMI-LINUX-1_CA20_NNP 1.00E+00
CA-20 - CiscoSW-1 "VALVE CLOSE" Signal to Opto22-2 Not Provided when Needed.
CISCOSW-1_CA-20_NNP 1.00E+00
CA21 - HMI-LINUX-1 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
HMI-LINUX-1_CA21_NNP 1.00E+00
CA-21 - CiscoSW-1 "TRIP VALVE CLOSE" Signal to Opto22-2 Not Provided when Needed.
CISCOSW-1_CA-21_NNP 1.00E+00
Valve Fails to Close
VALVE-FTC 1.00E-04
Valve Fails Open
VALVE-FO 1.00E-06
Valve Positioner Fails
VALVE-POS 1.00E-06
Loss of Power for Valve
VALVE-LOP 1.00E-06
Pump Fails to Start/Run
HVAC-CTDD-STPA-PUMP
PF40-1 Does Not Send "PUMP ON" Signal to Pump
PF40-1_CA_ON
Opto22-1 Does Not Send "PUMP ON" Signal to PF40-1
OPTO22-1_CA_ON
CA3 - PF40-1 "PUMP ON" Signal Not Provided when Needed.
PF40-1_CA3_NNP 1.00E+00
HMI-LINUX-1 Does Not Send "PUMP ON" Signal to Opto22-1
HMI-LINUX-1_CA_ON
CiscoSW-1 Does Not Send "PUMP ON" Signal
CISCOSW_CA_ON
CA1 - Opto22-1 "PUMP ON" Signal Not Provided when Needed.
OPTO22-1_CA1_NNP 1.00E+00
CA15 - HMI-LINUX-1 "PUMP ON" Signal Not Provided when Needed.
HMI-LINUX-1_CA15_NNP 1.00E+00
CA-15 - CiscoSW-1 "PUMP ON" Signal to Opto22-1 Not Provided when Needed.
CISCOSW-1_CA-15_NNP 1.00E+00
PF40-1 Sends "PUMP OFF" Signal to Pump
PF40-1_CA_OFF
Opto22-1 Sends "PUMP OFF" Signal to PF40-1
OPTO22-1_CA_OFF
CA4 - PF40-1 "PUMP OFF" Signal Provided when Not Needed.
PF40-1_CA4_PNN 1.00E+00
HMI-LINUX-1 Sends "Pump OFF" Signal to Opto22-1
HMI-LINUX-1_CA_OFF
CiscoSW-1 Sends "PUMP OFF" Signal to Opto22-1
CISCOSW_CA_OFF
CA2 - Opto22-1 "PUMP OFF" Signal Provided when Not Needed.
OPTO22-1_CA2_PNN 1.00E+00
CA16 - HMI-LINUX-1 "PUMP OFF" Signal Provided when Not Needed.
HMI-LINUX-1_CA16_PNN 1.00E+00
CA-16 - CiscoSW-1 "PUMP OFF" Signal to Opto22-1 Provided when Not Needed.
CISCOSW-1_CA-16_PNN 1.00E+00
Pump Fails to Start
PUMP-FTS 1.00E-03
Pump Fails to Run
PUMP-FTR 1.00E-03
Loss of Power for Pump
PUMP-LOP 1.00E-06

Figure 5. SIFT which is inclusive of traditional basic events & hazardous control actions identified in Table 4 4 for HAZCADS analysis of "HVAC cooling tank drain down" hazard for the LabVolt System.

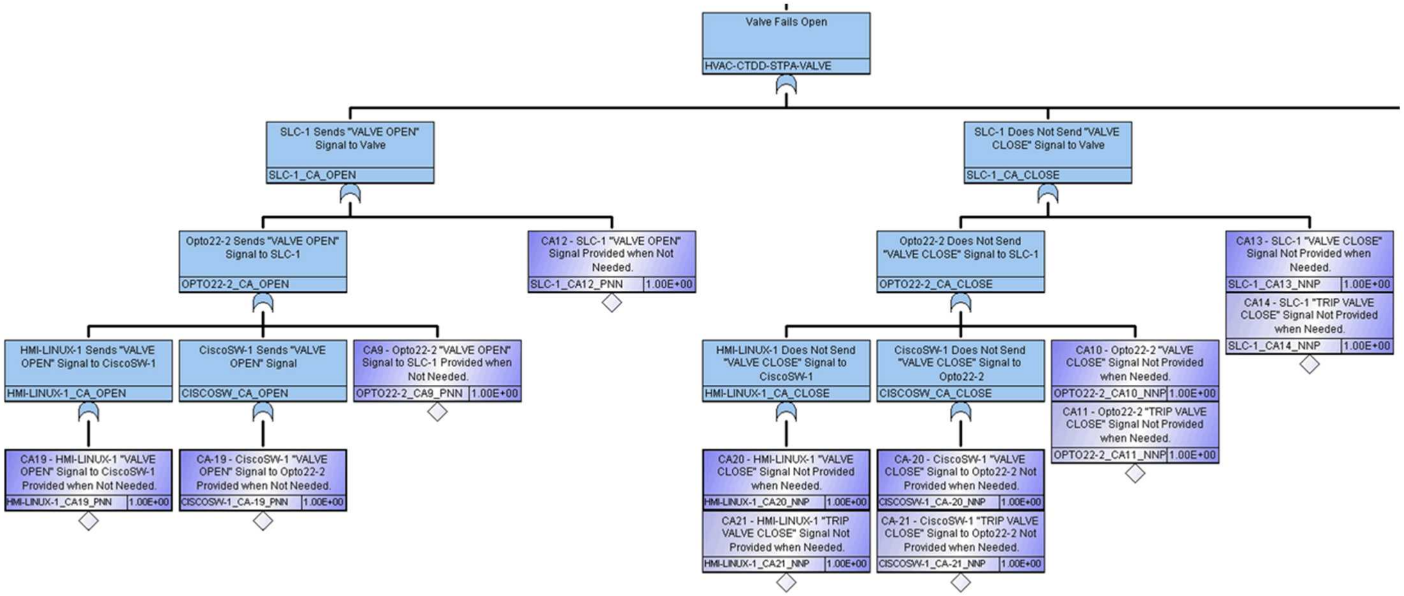
Valve Fails Open
HVAC-CTDD-STPA-VALVE
SLC-1 Sends "VALVE OPEN" Signal to Valve
SLC-1_CA_OPEN
SLC-1 Does Not Send "VALVE CLOSE" Signal to Valve
SLC-1_CA_CLOSE
Opto22-2 Sends "VALVE OPEN" Signal to SLC-1
OPTO22-2_CA_OPEN
CA12 - SLC-1 "VALVE OPEN" Signal Provided when Not Needed.
SLC-1_CA12_PNN 1.00E+00
Opto22-2 Does Not Send "VALVE CLOSE" Signal to SLC-1
OPTO22-2_CA_CLOSE
CA13 - SLC-1 "VALVE CLOSE" Signal Not Provided when Needed.
SLC-1_CA13_NNP 1.00E+00
CA14 - SLC-1 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
SLC-1_CA14_NNP 1.00E+00
HMI-LINUX-1 Sends "VALVE OPEN" Signal to CiscoSW-1
HMI-LINUX-1_CA_OPEN
CiscoSW-1 Sends "VALVE OPEN" Signal
CISCOSW_CA_OPEN
CA9 - Opto22-2 "VALVE OPEN" Signal to SLC-1 Provided when Not Needed.
OPTO22-2_CA9_PNN 1.00E+00
HMI-LINUX-1 Does Not Send "VALVE CLOSE" Signal to CiscoSW-1
HMI-LINUX-1_CA_CLOSE
CiscoSW-1 Does Not Send "VALVE CLOSE" Signal to Opto22-2
CISCOSW_CA_CLOSE
CA10 - Opto22-2 "VALVE CLOSE" Signal Not Provided when Needed.
OPTO22-2_CA10_NNP 1.00E+00
CA11 - Opto22-2 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
OPTO22-2_CA11_NNP 1.00E+00
CA19 - HMI-LINUX-1 "VALVE OPEN" Signal to CiscoSW-1 Provided when Not Needed.
HMI-LINUX-1_CA19_PNN 1.00E+00
CA-19 - CiscoSW-1 "VALVE OPEN" Signal to Opto22-2 Provided when Not Needed.
CISCOSW-1_CA-19_PNN 1.00E+00
CA20 - HMI-LINUX-1 "VALVE CLOSE" Signal Not Provided when Needed.
HMI-LINUX-1_CA20_NNP 1.00E+00
CA-20 - CiscoSW-1 "VALVE CLOSE" Signal to Opto22-2 Not Provided when Needed.
CISCOSW-1_CA-20_NNP 1.00E+00
CA21 - HMI-LINUX-1 "TRIP VALVE CLOSE" Signal Not Provided when Needed.
HMI-LINUX-1_CA21_NNP 1.00E+00
CA-21 - CiscoSW-1 "TRIP VALVE CLOSE" Signal to Opto22-2 Not Provided when Needed.
CISCOSW-1_CA-21_NNP 1.00E+00

Figure 6. Zoomed in portion of the lower gate Valve Fails Open (portion within the red brackets) from the SIFT in Figure 4 6 for HAZCADS analysis of "HVAC cooling tank drain down" hazard for the LabVolt System.

Table 1. Representative examples of the three types of cut sets from evaluating the *HVAC Cooling Tank Drain Down* event from SIFT analysis of the LabVolt System.

Cut Set ID #	SIFT Cut Set Type	Total # of Cuts Sets [% of total]	Representative Event Label & Description
1	Type 3	96 [~54%]	*OPTO22-2_CA9_PNN*: CA9 - Opto22-2 "VALVE OPEN" Signal to SLC-1 Provided when Not Needed.
4	Type 2	68 [~39%]	*OPTO22-2_CA9_PNN*: CA9 - Opto22-2 "VALVE OPEN" Signal to SLC-1 Provided when Not Needed.
7	Type 1	12 [~7%]	*PUMP-FTS*: Pump Fails to Start

Results further show that 120 cut sets were comprised *only* of combinations of hazardous control actions are associated with the digital components that make up the LabVolt system. In other words, solving SIFTs shows that system hazards (or fault tree top events) can be achieved *entirely* from digital component interactions; namely, hazardous control actions related to digital assets. In addition, of the 120 Type 3 cut sets, 18 are combinations of hazardous control actions from a single digital component: either the HMI-LINUX-1 or the CISCOSW-1. Considering that the traditional FTA cut sets and a majority of the SIFT cut sets identified that *two* events were necessary to achieve a system hazard, this is an important insight. More specifically, the HMI-LINUX-1 and the CISCOSW-1 in this LabVolt example can each initiate combinations of non-traditional digital component interactions with the potential to create a hazardous system state. These two components represent a single point digital component vulnerability in the LabVolt DI&C system. Lastly, solving SIFTS can also identify digital components that *do not* contribute to the hazardous system state. For example, the fan controller (SLC-2) in the HCS is not required to maintain adequate HVAC cooling tank water level. If HVAC cooling tank drain down was the only hazard of concern to system personnel in this example, then the fan controller can be omitted from subsequent cyber hazard analyses.

Insights from applying SIFTs to the LabVolt system indicate that a digital component hazardous control action to the pump *can cause* nontraditional pump faults due to the interactions of the pump and pump controller that drives the system towards a hazardous state.

HAZCADS OVERVIEW

The ultimate result of this EPRI-sponsored research was the development of the *Hazards and Consequences Analysis for Digital Systems* (HAZCADS) analysis technique.[9] HAZCADS uses SIFTs to efficiently and methodically address hazards and consequences that can emerge from digital systems. In addition, this analysis technique incorporates outcomes from each without significantly altering either STPA or FTA, as shown in Figure 7. The reporting in [9] offers more details for each HAZCADS activity, but the HAZCADS analytical process is summarized below:

- **HAZCADS Activity 3.1**: Identify plant operations, P&ID, DI&C logic models, and hazard-related documents to understand digital components and physical process relationships;
- **HAZCADS Activity 3.2**: Use STPA to identify system hazards and losses to construct a hierarchical control structure (HCS) that captures digital/physical/human interactions;
- **HAZCADS Activity 3.3**: Use STPA to evaluate digital controllers to identify possible opportunities for unsafe control actions to drive system behaviors towards hazardous states;

- **HAZCADS Activity 3.4**: Use FTA to identify the combination(s) of critical process (including digital) component failures that cause the top-level event to occur;
- **HAZCADS Activity 3.5**: Incorporate STPA-derived unsafe control actions into fault tree models to create SIFTs, identify prevention sets from SIFTs, and evaluate component importance; and,
- **HAZCADS Activity 3.6**: Determine the control method effectiveness score based on prevention set and importance analysis.

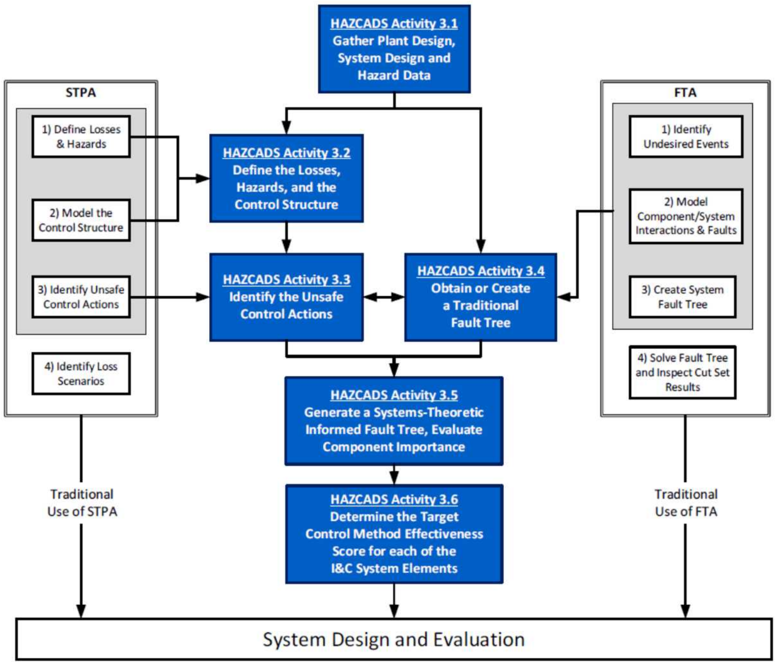
HAZCADS Activity 3.1
Gather Plant Design, System Design and Hazard Data
STPA
1) Define Losses & Hazards
2) Model the Control Structure
3) Identify Unsafe Control Actions
4) Identify Loss Scenarios
HAZCADS Activity 3.2
Define the Losses, Hazards, and the Control Structure
HAZCADS Activity 3.3
Identify the Unsafe Control Actions
HAZCADS Activity 3.4
Obtain or Create a Traditional Fault Tree
FTA
1) Identify Undesired Events
2) Model Component/System Interactions & Faults
3) Create System Fault Tree
4) Solve Fault Tree and Inspect Cut Set Results
HAZCADS Activity 3.5
Generate a Systems-Theoretic Informed Fault Tree, Evaluate Component Importance
HAZCADS Activity 3.6
Determine the Target Control Method Effectiveness Score for each of the I&C System Elements
Traditional Use of STPA
Traditional Use of FTA
System Design and Evaluation

Figure 7. Graphical representation of HAZCADS (blue) that identifies where analytical aspects of STPA and FTA are incorporated (EPRI 2018).

HAZCADS leverages the ability of STPA to represent potential DI&C failure modes that can be incorporated into fault trees to enable a comprehensive overview of functional responsibilities and hierarchical relationships in complex systems. Similarly, the primary outputs from FTA used in HAZCADS are the fault tree models. HAZCADS inserts the unsafe control actions generated using the early steps of STPA as undeveloped events[3] into traditional fault tree models. Thus, *digital* system faults which correspond to the non-traditional failure modes—or unsafe control actions—identified in HAZCADS must be aligned in the fault tree where component failures are identified.

As previously demonstrated, SIFTs are solved using the same Boolean algebraic logic that has been the foundation for FTA for decades, with the resultant cut sets evaluated qualitatively. As a result, HAZCADS can be applied across all the safety and non-safety systems in a nuclear power plant and will result in cut sets describing a range of potential failure pathways across the complex system. As additional benefit is that HAZCADS can be applied to many different losses (e.g., core damage, plant trip, loss of emergency preparedness systems).

[3] Modeling STPA unsafe control actions as undeveloped events give awareness to the fact that unsafe control actions still need to be fully comprehended (e.g., through causal analysis).

CONCLUSIONS & IMPLICATIONS

The analytic utility of SIFTs has helped increase interest in—and development of—HAZCADS. For example, these outcomes have prompted EPRI to formally include HAZCADS into its current strategy for improving cyber security within the U.S. domestic NPP fleet. There have also been preliminary discussions with EPRI on wider application of HAZCADS, including beyond cyber security-related applications, such as its use in EPRI's Digital Engineering Guide [10] that addresses digital reliability. In addition, Sandia researchers have assisted in developing training materials for the methodology's use in the U.S. domestic fleet of nuclear power plants and has supported several pilot implementation projects in early 2019. These pilot applications of HAZCADS by U.S. nuclear power plants resulted in users identifying digital hazards that caused a real nuclear power plant trip scenario. This is impactful given that the current leading causes of plant trips—which can result in operational cost losses of ~$1M a day—are related digital I&C.

In many ways, SIFTS (in general)—and HAZCADS (more specifically)—seem positioned to support both R&D efforts to improve risk-informed analysis of complex highly digital systems *and* bolster the development of industry guidance to better manage cyber-related risks at nuclear facilities.

REFERENCES

[1] National Research Council, 1997, *Digital Instrumentation and Control Systems in Nuclear Power Plants: Safety and Reliability Issues*. Washington, DC: The National Academies Press.

[2] Nuclear Energy Institute (NEI), 2010, *Cyber Security Plan for Nuclear Power Reactors*, NEI 08-09.

[3] U.S. Nuclear Regulatory Commission (USNRC), 1975, *Reactor Safety Study, An Assessment of Accident Risks in Nuclear Power Plants*, WASH-1400, NUREG/75-014.

[4] Electric Power Research Institute (EPRI), 2015(a), *Analysis of Hazard Models for Cyber Security – Phase I*, 3002004995, Palo Alto, CA.

[5] Electric Power Research Institute (EPRI), 2015(b), *Program on Technology Innovation: Cyber Hazards Analysis Risk Methodology – Phase II: A Risk Informed Approach*, 3002004997, Palo Alto, CA.

[6] Electric Power Research Institute (EPRI), 2016, *Cyber Security Technical Assessment Methodol-ogy – Vulnerability Identification and Mitigation*, 3002008023, Palo Alto, CA.

[7] U.S. Nuclear Regulatory Commission (USNRC), 1981, *Fault Tree Handbook*, NUREG-0492.

[8] Leveson, N., 2011, Engineering a Safer World – Systems Thinking Applied to Safety, MIT Press, Cambridge, MA.

[9] Electric Power Research Institute (EPRI), 2018, *Hazards and Consequences Analysis for Digital Systems*, 3002012755, Palo Alto, CA.

[10] Electric Power Research Institute (EPRI), 2016, *Digital Engineering Guide: Decision Making Using Systems Engineering*, 3002011816, Palo Alto, CA.