A CONTROL-THEORETIC APPROACH TO THE RESILIENT DESIGN OF EXTRA-TERRESTRIAL HABITATS

by

Robert Kitching

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Aeronautics and Astronautics

School of Aeronautics and Astronautics

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Karen Marais, Chair
School of Aeronautics and Astronautics

Dr. Shirley Dyke
School of Mechanical Engineering

Dr. Daniel DeLaurentis
School of Aeronautics and Astronautics

Approved by:
Dr. Gregory Blaisdell

Dedicated to my mother, Angela Kitching

ACKNOWLEDGMENTS

This material is based upon work supported in part by NASA under grant or cooperative agreement award number 80NSSC19K1076. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration (NASA).

I would like to thank my mother who has sacrificed so much for me to be able to be in the position that I am today and continues to humble me with her unrelenting support, and my partner Sarah Hirschman who has always supported me regardless of the physical distance throughout our higher education journeys.

I would like to thank the people that made this work possible. I want to thank the RETHi team, especially Dr. Amin Maghareh and Dr. Shirley Dyke, who continued to challenge me throughout my involvement in the project; Hunter Mattingly and Dale Williams, who were integral in the creation of the database; all the students involved in the two semesters of our space systems safety course whose enthusiasm in research helped populate the database; and Dr. Daniel DeLaurentis without whom I would not have had the idea to create a network.

Thank you to everyone in the VRSS research group whose efforts in research helped push me to be the best researcher I could be, and to all the students I have met in and outside of Armstrong Hall whose passion for Aerospace reflects the incredible culture at Purdue University.

Finally, to my adviser, Dr. Karen Marais: thank you for seeing something in me as an undergraduate and giving me the chance to join your research group and be a part of the impressive research culture that you have created. Thank you for supporting me throughout my graduate school journey, and for helping me become a better communicator, writer, presenter, and person.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Space habitats will involve a complex and tightly coupled combination of hardware, software, and humans, while operating in challenging environments that pose many risks, both known and unknown. It will not be possible to design habitats that are immune to failure, nor will it be possible to foresee all possible failures. Rather than aiming for designs where "failure is not an option", habitats must be resilient to disruptions. We propose a control-theoretic approach to resilient design for space habitats based on the concept of safety controls from system safety engineering. We model disruptions using a state and trigger model, where the space habitat is in one of three distinct states at each time instance: nominal, hazardous, or accident. The habitat transitions from a nominal state to hazardous states via disruptions, and further to hazardous and accident states via triggers. We develop an approach for identifying safety controls that considers these disruptions, hazardous states, and identifies control principles and their possible control flaws. We use safety controls as ways of preventing a system from entering or remaining in a hazardous or accident state. We develop a safety control option space for the habitat, from which designers can select the set of safety controls that best meet resilience, performance, and other system goals. We show how our approach for identifying safety controls drives our control-theoretic approach for resilient design, and how that fits into the larger system safety engineering process. To identify and assess hazards, we use a database and create a network format that stores the relationships between different disruptions and hazardous states for an example space habitat. We use this database in combination with traditional hazard assessment techniques to prioritize control of possible disruptions and hazardous states. To mitigate hazards, we develop a safety control option space that contains safety controls that either prevent transition to hazardous states or return the habitat to a nominal state. We use generic safety controls, or the principle of control, to generate new safety controls as our set of disruptions and hazardous states grows, and store these in the database. Lastly, we evaluate our mitigation techniques using our control effectiveness metric, a metric intended to assess how well a safety control addresses the hazardous state or disruption that it is designed for. Our control-theoretic approach is one way in which we can complete the system safety engineering process for a space habitat system and can provide design guidance for the development of resilient space habitats.

1. INTRODUCTION

1.1 Motivation

Space habitats will involve a complex and tightly coupled combination of hardware, software, and humans. These habitats will be embedded in challenging environments, whether the microgravity of cislunar space, or the surface of the Moon or Mars. These harsh environments pose many risks, both known and unknown. In an extraterrestrial environment, it is inevitable that things will go wrong. Failures and faults may include components failing, operator or software implementing the wrong actions, or dysfunctional interactions between correctly functioning components. Space habitat systems will be safety-critical, meaning that a "failure might endanger human life, lead to substantial economic loss, or cause extensive environmental damage" (Knight, 2002). It will not be possible to design habitats that are immune to failure, nor will it be possible to foresee all failures. Therefore, rather than aiming for designs where "failure is not an option", we must design habitats that are resilient to the inevitable failures that will occur. Resilience is the ability of a system, process, or organization to react to, survive, and recover from disruptions (Uday & Marais, 2015). Designing for resilience ensures that a system can *adapt* before or during an encounter with a threat, *prepare for* a threat in advance to enable recovery following an encounter, *withstand* a threat by retaining partial or full functionality following an encounter, or *recover from* a threat by restoring partial or full functionality following an encounter.

1.1.1 Resilience

Almost all complex systems are designed to be reliable, in that they are designed with the ability to deal with known and credible threats and continue to function despite these known threats (Panteli, 2015). A new performance marker, resilience, has made it more evident that more needs to be considered when evaluating how systems deal with threats or disruptions. The U.K. Cabinet Office claims resilience "encompasses reliability and it further includes resistance, redundancy, response, and recovery as key features" (Panteli, 2015). Resilience is often referred to as a system's ability to quickly and effectively "bounce back" from a disturbance or interruption to its nominal performance. Resilience in engineering systems has been defined in many ways. All definitions have the elements of preparing for, surviving, and recovering from disruptions, as shown in Figure 1.1. At the time of disruption, the system withstands the impact a certain amount (labeled

"Surviving the disruption"), and the system recovers over time to its nominal performance level (labeled "Recovering from the disruption"). The ability of a system to maintain functionality can be identified as "static resilience", while the recovery of the system after a disruption can be identified as "dynamic resilience" (Rose, 2007). Maximizing resilience includes minimizing system degradation from a disruption and decreasing the time from the system's performance level at the time of disruption to its regained performance level. In this sense, we view system resilience as a multifaceted notion, represented as a combination of survivability and recoverability. This conceptualization is widely used in the literature to depict the fundamental ideas behind resilience (Uday & Marais, 2015).

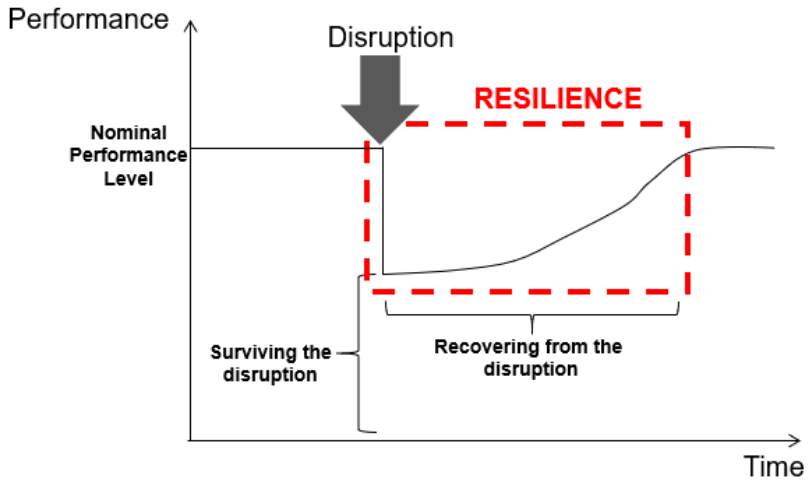

Figure 1.1: The Resilience Curve

Resilience is dependent on context, and it depends on the structure or architecture of the system (Uday & Marais, 2015) causing variation in possible resilience curves. A disruption may not cause the system performance to drop deeply and suddenly, but rather a gradual decline may be observed. For example, in terms of space habitat internal pressure capabilities, a micrometeoroid breaching the habitat structure may degrade the performance of the habitat significantly and suddenly, but a small pressure leak would be observed as a gradual decline. Similarly, there are different methods for a system to recover from disruptions. Measures to ensure recovery may include an increase in performance after a recovery to make up for lost capability. Conversely, disruptions can have long-term impacts on systems. Continuing our example, repairing habitat structure with material with inferior adhesive capabilities may have long term effects on the pressure of the habitat. But,

modifying the habitat environmental control system to regulate the pressure at a higher capability to offset the lost pressure would mean an increase in habitat performance.

Jackson & Ferris (2012) developed a set of resilience principles for engineered systems that are grouped into four resilience attributes. These attributes are “capacity, or the ability of the system to survive a threat; flexibility, or the ability of the system to adapt to the threat; tolerance, or the ability of the system to degrade gracefully in the face of a threat, and cohesion; or the ability of the system to act as a unified whole in the face of a threat”. Our resilient design approach needs to incorporate these attributes so that the space habitat can adequately handle and respond to the threats present in an extraterrestrial environment. The habitat design must be robust in that it will be capable of absorbing disruptions and keep operating. The design must also be able to manage a disruption as it unfolds, and the habitat must have the ability to recover rapidly after a disruption.

1.1.2 Risk Analysis Techniques

Resilience has emerged as a research topic within the last twenty years (Patriarca et al., 2018), but system safety engineering, including the techniques used in the field, has been around much longer. Many of the techniques rely on assessments of component reliability and are not properly equipped to handle the inherent interdependencies that resilient design must account for in tightly coupled and complex systems. All risk managers must ask four key questions: 1) what can go wrong (and how)? 2) how likely is it? 3) what are the consequences? and 4) what can be done about it? To sufficiently address question 4), me must first adequately complete question 1).

An established method of understanding and assessing what can go wrong, or conducting hazard identification and assessment, is to use conventional reliability design approaches such as a hazard and operability study (HAZOP) or Preliminary Hazard Analysis (PHA). Conventional binary and event-based safety and reliability design approaches cannot respond adequately to the level of complexity found and rapid failure responses needed in space habitat systems. Traditional, component-centric approaches to risk identification, assessment, and management, manage risk by preventing failures, or reducing the effects of failures. The weaknesses and limitations of these approaches have been well-documented, and center about their inability to properly address software, human interactions, and accidents that do not involve component failure, but rather arise because of dysfunctional interactions. The more complex the system, the less applicable

assumptions like independence of failures become. When we include the potential for different environmental conditions, not all of which may be foreseen either in terms of their type (e.g., physical/chemical characteristics of particulates in the atmosphere) or extent (e.g., frequency/intensity of storms), identifying and assessing risk becomes even more challenging. Most approaches to hazard identification (e.g., HAZOP) are essentially sophisticated checklists. In complex unprecedented systems hazard identification is especially difficult because many of the hazards are unprecedented or cannot be foreseen. While failure modes, effects, (and criticality) analysis (FMEA/FMECA) may also help with hazard identification, it requires each failure to be considered on an individual basis, and each failure is considered independently. Event Trees and Event Sequence Diagrams are also inductive techniques where a basic initiating event is propagated to its potential consequences. The analyst must know which components or initiating events to consider – in a complex system like a space habitat it is infeasible to analyze all of them (Leveson et al., 2009). Finally, these techniques do not allow consideration of interconnected or otherwise dependent failures.

1.1.3 Accident Modeling Techniques

Using accident models is another way to understand what can go wrong when conducting a risk assessment. Accident models are important to system safety engineering because they underlie all efforts to engineer for safety. They explain why accidents occur, they determine the way to prevent and investigate accidents, and they impose patterns on accidents (Leveson, 2004). Most accident models view accidents in the form of a chain or sequence of events. Event-based models imply direct causality, especially linear causality. It is difficult to include non-linear relationships and component interactivity. These chain-of-events models are limited by assumptions (Leveson, 2011). These assumptions are that "1) decomposition of the system so that physical aspects are decomposed into separate physical components while behavior is decomposed to events over time assumes that such separation is feasible, 2) the components or events are not subject to feedback loops and other non-linear interactions and that the behavior of the components is the same when examined singly as when they are playing their part in the whole, and 3) the interactions among the subsystems are simple enough that they can be considered separate from the behavior of the subsystems themselves". These assumptions are reasonable for many simpler systems, and these

models are effective for losses caused by hardware failure and for non-complex systems, but for the complexity we are considering new approaches are needed.

1.2 Resilient Extra-Terrestrial Habitats institute (RETHi)

The work in this thesis is work done in conjunction with the Resilient ExtraTerrestrial Habitat institute (RETHi), a NASA sponsored Space Technology Research Institute hosted at Purdue University and including representatives from the University of Connecticut, Harvard University, and the University of Texas at San Antonio. The vision of the institute is to "develop and demonstrate transformative smart autonomous habitats and related technologies that will adapt, absorb, and rapidly recover from expected and unexpected disruptions to deep space habitat systems without fundamental changes in function or sacrifices in safety". The institute is divided into three research thrusts. First, the system resilience thrust aims to develop techniques needed to establish a control-theoretic paradigm for resilience, and the computational capabilities needed to capture complex behaviors and perform trade studies to weigh different choices regarding habitat architecture and onboard decisions. Second, the situational awareness thrust aims to develop and validate generic, robust, and scalable methods for detection and diagnosis of anticipated and unanticipated faults that incorporates an automated active learning framework with robots and humans in the loop. Third, the robotic maintenance thrust aims to develop and demonstrate the technologies needed to realize teams of independent autonomous robots, that navigate through dynamic environments and perform tasks such as collaboratively replacing damaged structural elements. This thesis falls under the system resilience thrust, and applies various assumptions based on the other research thrusts. For example, we assume that we have the capabilities to detect and diagnose faults in our system based on the situational awareness thrust. Also, we assume that there are robots capable of autonomously performing repairs and a variety of other tasks based on the robotic maintenance thrust.

The objective of this work in the context of the RETH institute is to advance the control-theoretic approach to resilience that supports smart habitat system architecture. We leverage research in system safety engineering and accident modeling, use lessons learned from previous incidents and accidents in space, and build on previous work in risk analysis and resilience design to propose a new approach to resilient design for space habitats based on using a control-theoretic view of

resilience for space habitats to maintain the habitat in question within a safe boundary of system performance.

2. CONTROL-THEORETIC APPROACH TO RISK ANALYSIS

What is needed for resilient space habitats is an approach that (1) goes beyond the event-centric failure model underlying conventional risk-based design, and (2) helps identify designs that are prepared for both foreseen and unforeseen risks. To address (1), we propose an approach that uses hazardous states and triggers into these states as its basic elements, rather than component and other failures. We then use safety controls to prevent transition to hazardous states, or to allow exit from hazardous states. This thesis forms part of a larger research effort, RETHi, that intends to apply this control-theoretic approach to create and investigate resilient space habitat architectures. The RETHi project spans five years, and in that time the institute aims to demonstrate methods and tools that support resilient space habitat design. Some of these methods include a computational platform that models the habitat systems, robot and human agents, and a health management system, a cyber-physical testbed that integrates these computational models into a physical habitat structure, and most importantly for our applications, the concept of a resilience power metric to address (2), and assess how well safety controls address unforeseen disruptions or hazardous states and contribute to overall habitat resilience.

We consider safety as a control problem, where safety is an emergent property of the system. Rasmussen (1997) pioneered the effort to use control theory in accident modeling: he argued that accidents tend to be caused by a systematic migration of organizational behavior to the boundaries of safe behavior caused by pressures relating to cost-effectiveness and a competitive environment, and not by a coincidence of independent failures. The concept of "boundaries of safe behavior" introduces the conceptualization of regions of safe behavior of the system and regions of unsafe behavior of the system. Rather than assessing faults and failures and reducing their effects, control-theoretic approaches assess risk based on how well the system is kept within safe operating states, or conversely, how well it is kept out of unsafe, or hazardous states (Leveson et al., 2009). Humans and organizations can adapt to foreseen and unforeseen threats and still maintain safety if they stay out of regions of unsafe behavior (Leveson, 2004). Leveson's (2004) Systems-Theoretic Accident Model and Processes (STAMP) model uses systems theory to show that accidents occur when disturbances, failures, or dysfunctional interactions among system components are inadequately controlled by safety-related constraints on the development, design, and operation of the system.

By moving away from the component-centric view of risk, these approaches account for all types of accidents, including those that arise without any components failing.

2.1 The Control-Theoretic Approach Steps

To demonstrate our control-theoretic approach to risk analysis, we define several terms and specify four steps that we use to mitigate risks, and to keep our system operating in a region of safe behavior. Figure 2.1 shows a visual representation of our control-theoretic approach.

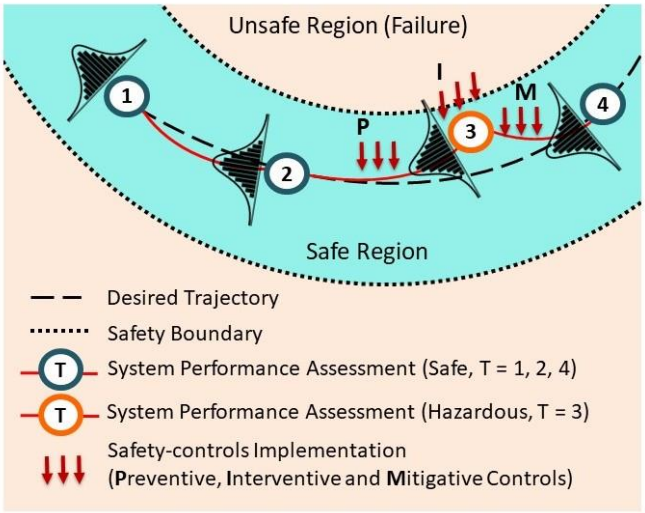

Figure 2.1: Visual Representation of the Control-Theoretic Approach to Risk Analysis [RETHi, 2020]

From Figure 2.1, we define steps to achieve our control-theoretic approach and define terms that we will use in the next section in our state and trigger model.

Step 1 is to identify events, or *disruptions*, that could cause the system to propagate from a region of safe behavior and enter a region of unsafe behavior. The region of safe behavior is what we call the *nominal state*, in that the system is operating as normal. A disruption to the system causes the system to transition to a region of unsafe behavior, or a *hazardous state*. Also involved in step 1 is the identification and definition of the hazardous states that occur as a result of these disruptions.

Step 2 is to assess these disruptions and hazardous states. A single disruption may cause multiple hazardous states, and multiple disruptions may cause the same hazardous state. In this step, we assess the impact of each disruption and hazardous state on the system.

Step 3 is to implement actions or design choices, or *safety controls*, that return the habitat to the nominal state or prevent behavior that will result in an accident or loss, or an *accident state*.

Step 4 is to assess the effectiveness of each safety control of controlling the system or keeping the system within a region of safe behavior. We achieve this step by creating a *control effectiveness* metric, a metric that is intended to indicate how well a certain safety control addresses the hazardous state or disruption it was designed for. The remainder of this thesis follows these steps, as Chapter 3 covers step 1, Chapter 4 covers step 2, Chapter 5 covers step 3, and Chapter 6 covers step 4.

2.1.1 The State and Trigger Model

Our approach considers systems as being in one of three types of states: nominal, hazardous, or accident. A state is a segment of time in which a system exhibits a certain behavior. A system can be in one and only one state at a given time. A nominal state is when the system is within the boundaries of safe behavior. A hazardous state is when the system is in a state that, if left uncontrolled, will result in an accident or loss of life. Triggers transition the system from one state to another. Triggers instantiated events and cause a system to transition between states or remain in the same state (Rao & Marais, 2020). Each state must have at least one entering trigger. Disruptions are a type of trigger that instigates transition to a hazardous or accident state.

We use safety controls to maintain the system in nominal states, or, if it does transition to a hazardous or accident state, return it to a nominal state. A safety control is any part of the system design or operation that maintains the system in a nominal state, prevents the system from propagating to a hazardous state, or restores the system from a hazardous or accident state to a nominal state.

We use safety controls at the time of the disruption to prevent the system from entering a hazardous state. Or, we can use safety controls after the time of the disruption to prevent the system from entering a hazardous state or regain the system performance to a nominal state. Figure 2.2 shows

an example of a simple system with three possible states modeled as a state and trigger model. In this example, we are concerned with a possible pressure loss inside the habitat due to a micrometeoroid impact. The system begins in the nominal state. For convenience, we omit the implied entry trigger into the first nominal state. We model the micrometeoroid impact as an initiating disruption that could cause a breach in the habitat structure, triggering a transition to the hazardous state that the *habitat is losing pressure*. If no action is taken, the habitat may further deteriorate into a state of *unlivable pressure environment*. We identified several potential safety controls to prevent transition to the hazardous or accident states. Safety controls appear either as transition (to hazardous or accident state) preventers, as shown in the figure by the green crosses, or as triggers away from hazardous or accident states, as shown for example by the green trigger that returns the system to the nominal state from the hazardous state.

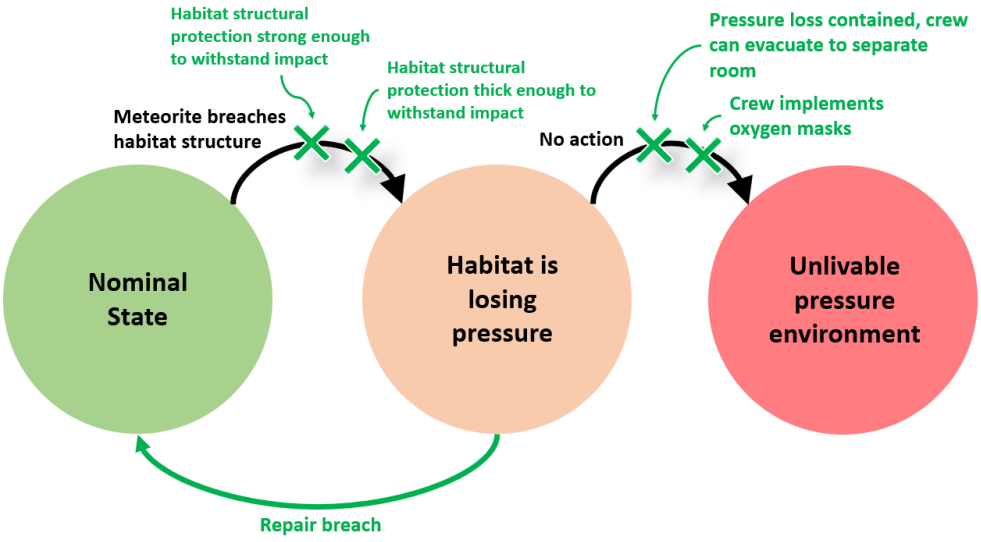

Figure 2.2: Example State and Trigger Model

2.2 Traditional System Safety Process

Bahr (2016) describes that the overall purpose of the system safety process is to identify hazards, eliminate or control them, and mitigate the residual risks.

The first step in the system safety process is to define the boundary conditions or analysis objectives. This is the scope or level of protection desired for the system. The designer should answer the question “How safe is safe enough?”, understand what constitutes critical accidents,

and understand the accepted cost of preventing accidents, among other things. This step involves defining criticality of accidents from catastrophic to negligible.

The next step is system description. In this step, we should understand how the system works and the interactions between the hardware, software, people, and environment. Describing the system accurately is essential to the safety analysis.

The next step is hazard identification, and the purpose is to identify all hazards that may affect the system. There are many methods to achieve this step, however it is at its core a safety brainstorming session. Once the hazards are identified, the next step is hazard analysis in which we study how each hazard affects the system. This step is often completed via a Preliminary Hazard Analysis or System Hazard Analysis, which we cover in Chapter 4. Once hazards have been identified and analyzed, the next step is to evaluate the hazard risks. Or, how likely each hazard is to occur, and if it does how much damage will result. These steps assist in sufficiently assessing which risks require control.

The next step is hazard control, in which we must control the effects of the hazards. We may design out hazards with engineering design changes or operational procedures, or we may use management controls to make changes to the organization itself (e.g., a production plant safety plan). Once controls are in place, the next step is to verify that the controls adequately control the identified hazards or mitigate the risks. Once that step is completed, the next step is to make the formal decision that the residual risk in the system is acceptable. If the risk is unacceptable, the system is then modified and the risks are re-assessed. If the risk is acceptable, the design changes are documented, and risk is periodically reviewed from that point onwards. In the next section, we map our control-theoretic approach to the established system safety process.

2.3 Mapping the Control-Theoretic Approach to the System Safety Process

Figure 2.3 shows how our control-theoretic approach fits into the overall system safety process.

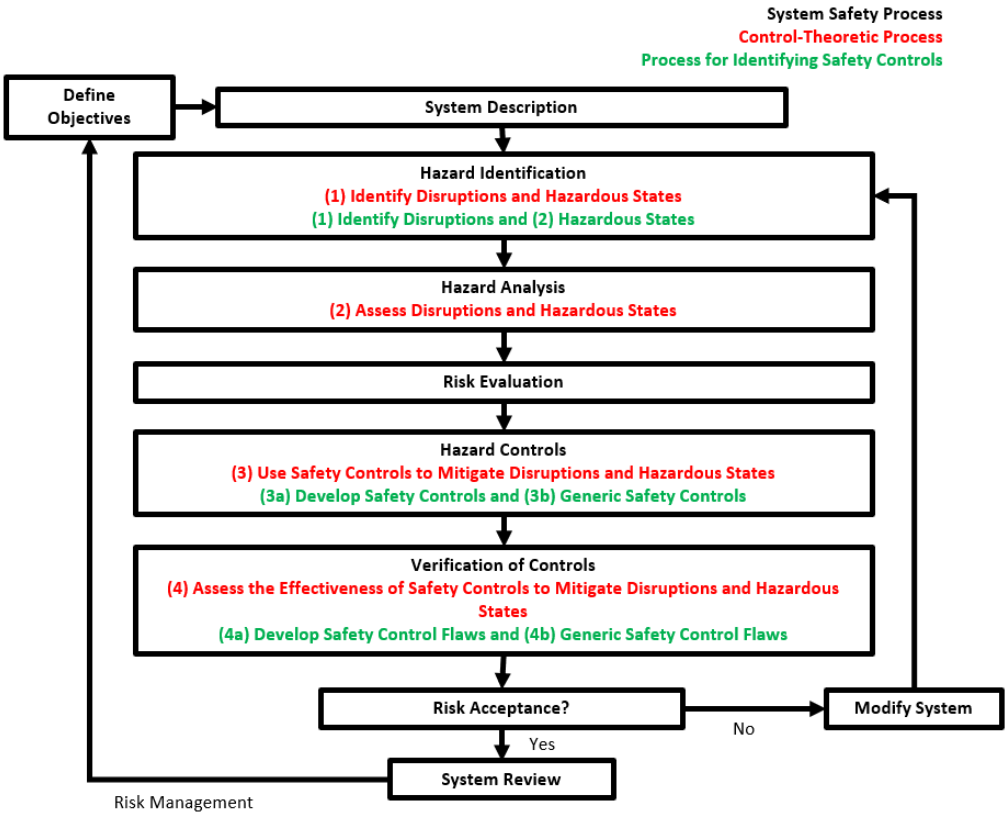

Figure 2.3: The Control-Theoretic Process and the Process for Identifying Safety Control Mapped to the System Safety Process

The remaining chapters of this thesis cover the four steps outlined in Section 2.1 and shown in Figure 2.3. The first step in the control-theoretic process is to identify disruptions and hazardous states that may be present in an extraterrestrial environment. Additionally, identifying disruptions is the first step in the process of identifying safety controls, and identifying the hazardous states that result from those disruptions is the second step in the process of identifying safety controls. These steps are covered in Chapter 3 and are all contained under the larger *Hazard Identification* step in the system safety process. Next, under the *Hazard Analysis* system safety process step, we complete the second step in the control-theoretic approach which is to assess the disruptions and hazardous states. This is covered in Chapter 4. The third step in our control-theoretic approach is to use safety controls to mitigate disruptions and hazardous states. To use safety controls, we must complete steps 3a and 3b in our process for identifying safety controls, which are to develop safety controls and generic safety controls. These processes are covered in Chapter 5 and fall under the larger system safety process step *Hazard Controls*. Finally, to complete the system safety process

step *Verification of Controls*, we must assess the effectiveness of our safety controls to mitigate the disruptions and hazardous states. We achieve this through our control effectiveness metric, covered in Chapter 6. To support this step, we complete steps 4a and 4b in our process for identifying safety controls, which are to develop safety control flaws and generic safety control flaws.

We show in this chapter that our control-theoretic process is one way of completing the traditional risk management steps of hazard identification, hazard assessment, development of mitigative controls, and subsequent assessment of those controls. We note that our process differs in that we specify the control of disruptions and hazardous states, and our theory stems from implementing controls that maintain the system in regions of safe behavior. We do not consider "hazards" or "risks", but rather we model the system as a set of states and consider events that trigger transition between those states. We also note that our process for identifying safety controls is iterative, contained inside an iterative control-theoretic approach to system safety. In this thesis, we demonstrate our method and work through an example of an extraterrestrial example, however in practice the disruptions, hazardous states, and safety controls must be continuously updated and reviewed as more information on the system becomes available.

2.4 Thesis Objective and Outline

In this thesis, we demonstrate how we apply our control-theoretic approach to resilient design to an example extraterrestrial habitat, and how we use the established system safety engineering process to identify and assess disruptions and hazardous states, and identify and assess our mitigation techniques using safety controls, generic safety controls, control flaws, generic control flaws, and the control effectiveness metric.

This thesis first describes our control-theoretic approach for the design of resilient space habitats based on maintaining the habitat within a safe boundary of system performance, then it describes how we applied and evaluated this approach for an example Martian surface habitat. In Chapter 2, we explained the steps we have taken to develop our approach to resilient habitat design, based on an accident model that views the habitat as being in safe or unsafe states rather than considering component and other failures. In Chapter 3, we describe how we identify and enumerate a set of disruptions and hazardous states for an example habitat. and introduce our database for storing our

disruptions and hazardous states. In Chapter 4, we cover how we use hazard analysis techniques and our database of disruptions and hazardous states to assess the hazards and prioritize controls and mitigation techniques. In Chapter 5, we describe how we identify safety controls for our disruptions and hazardous states, and how we develop our safety control option space and generic safety controls. In Chapter 6, we cover how we develop our control effectiveness metric, and assess the effectiveness of our safety controls in mitigating disruptions and hazardous states. Chapter 7 concludes the thesis.

3. IDENTIFYING DISRUPTIONS AND HAZARDOUS STATES

In this chapter, we cover the first step in our control-theoretic approach to resilient design: identifying disruptions and hazardous states. This step is part of the larger system safety engineering process and is a method of *Hazard Identification*. We also cover the process for identifying safety controls, which begins with identifying disruptions and hazardous states, shown in Figure 3.1. The process for identifying safety controls informs our control-theoretic approach in the identification of disruptions and hazardous states, mitigating these disruptions and hazardous states, and assessing the effectiveness of our safety controls. Throughout the rest of this thesis, we will reference the steps shown in Figure 3.1 as we work through our example.

Using the state and trigger model as a basis, we propose the approach shown in Figure 3.1 for identifying safety controls. In this chapter, we focus on steps 1 and 2, identifying disruptions and hazardous states.

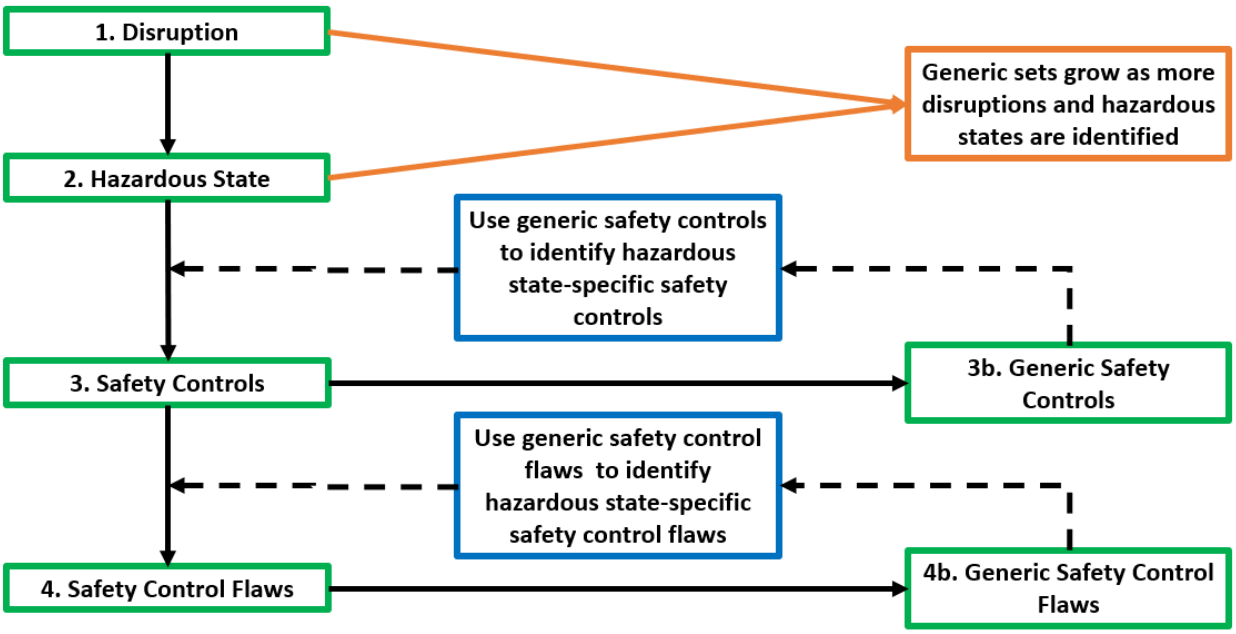

Figure 3.1: Approach for Identifying Safety Controls

To identify disruptions and hazardous states, we use established techniques from other fields (e.g., HAZOP) and domain expertise (e.g., NASA's lessons-learned database), to create a diverse (but

not necessarily complete) set of disruptions and hazardous states from which to begin. We add to this set of disruptions and hazardous states as the design matures and more information is available about the habitat.

To support the identification of extraterrestrial disruptions and hazardous states, we designed an undergraduate level course in the School of Aeronautics and Astronautics in which we conducted research on historical human space travel accidents and incidents and on space system safety engineering. We generated disruptions and hazardous states using a multitude of spaceflight incidents, used the state and trigger model to map these incidents, and stored the class findings in a database for use in this thesis. Additionally, we worked closely with two undergraduate aeronautical and astronautical engineering students to model specific relationships between possible disruptions and hazardous states that may occur. We investigated the constituent systems that we may expect to be present in a surface habitat, documented the interfaces between the systems, and tracked how example disruptions could affect each of these systems individually and affect the habitat as a whole. We also identified disruptions and hazardous states for the RETHi project during two case study exercises, where we worked closely with structural, mechanical, thermal, robotic, and power system engineers to better understand what may cause hazardous states, and what those hazardous states may be, in a habitat system. We leveraged expertise from RETHi, involved our undergraduate students, and investigated documented historical spaceflight events to develop a preliminary list of disruptions and their associated hazardous states. In the next section, we show how we identified these disruptions and hazardous states for an example Martian habitat.

3.1 Application: Identifying Disruptions and Hazardous States for a Martian Habitat

In this section we discuss identifying, linking, and recording disruptions and hazardous states for a space habitat and storing them in a database.

We consider for our case study a conceptual Mars surface habitat. The habitat consists of a group of connected domes in which the crew lives and performs day to day activities. The habitat has radiation and thermal protection, a photovoltaic power unit, an Environmental Control and Life Support System (ECLSS), and some autonomous and robotic capabilities. A further breakdown of the habitat systems is shown in Figure 3.2. The crew spend months at a time in the habitat due to

mission constraints on getting to and from Mars, and limited ground control support is available. The habitat must be designed to adapt to the harsh environment and be resilient against possible disruptions. Here we demonstrate Steps 1 and 2 of our approach.

Step 1: Identify Disruptions

We generated a preliminary list of disruptions by identifying the kinds of threats the habitat may encounter on Mars. Table 3.1 shows an excerpt of this list. For example, some of the more obvious disruptions we consider are a micrometeoroid impact to the habitat, ionizing radiation, and seismic activity in the area of the habitat. We also consider events like material outgassing, large variations of external temperature, and dust accumulation on the habitat.

Table 3.1: Preliminary List of Disruptions to Martian Habitat

Disruptions to Martian Habitat
High winds cause dust and debris to impact habitat
Ionizing radiation (including Galactic Cosmic Radiation)
Rapid rise in external temperature
Rapid decrease in external temperature
Extreme high external temperature
Extreme low external temperature
Outgassing of materials
Cold welding causes mechanical parts to fuse
Micrometeoroids impact habitat
Impact of ejecta
Seismic activity within/near habitat
Non-ionizing radiation
Dense dust surrounds habitat

Step 2: Identify Hazardous States

Next, we generate an initial list of possible hazardous states. To determine how disruptions affect the habitat and how the different functions of the habitat will respond to these disruptions, we break the habitat down into its constituent systems, as shown in Figure 3.2. For example, the structural system is composed of the parts of the habitat that contribute to the physical integrity of the habitat. The radiation protection system may include subsystems like multi-layered insulation or a regolith layer. Many of these systems are interconnected, such as the water recovery and management system providing water for the oxygen generation system through electrolysis, or more obviously the control system, which receives inputs from the sensor management system on habitat setpoints and works to distribute power to many of the other systems to maintain habitat functionality.

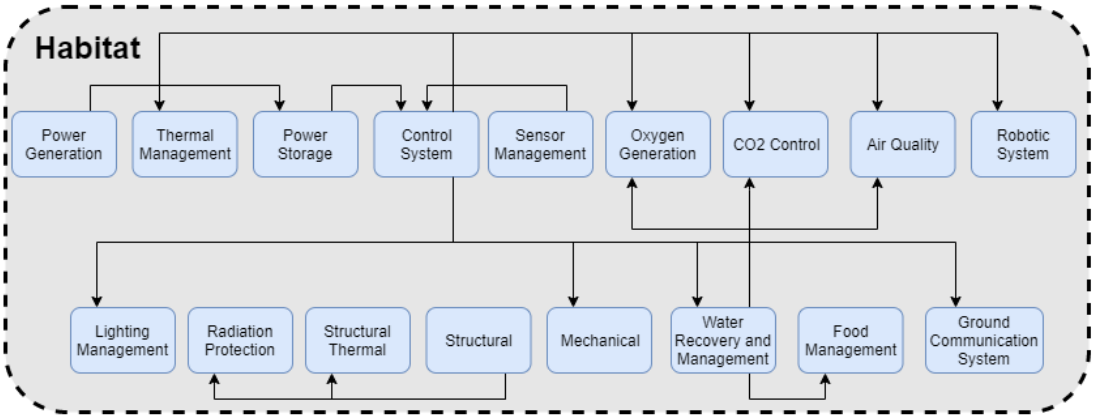

Figure 3.2: Habitat Systems and Resource Dependencies

This decomposition facilitates identifying and tracking how an initial disruption may propagate through the habitat. For example, Figure 3.3 shows how the effects of an initial micrometeoroid impact cascade through the habitat. The three hazardous states resulting from a micrometeoroid impact are of immediate concern and should be addressed through human or automated intervention. However, a performance loss in the habitat sensor management system could have more impactful habitat performance implications because of the number of systems that take inputs from the control system, which relies on the sensor management system. We use these system dependencies to record how a single disruption can cause multiple hazardous states in the system it directly impacts, and further cause hazardous states in systems that are dependent on other disrupted systems.

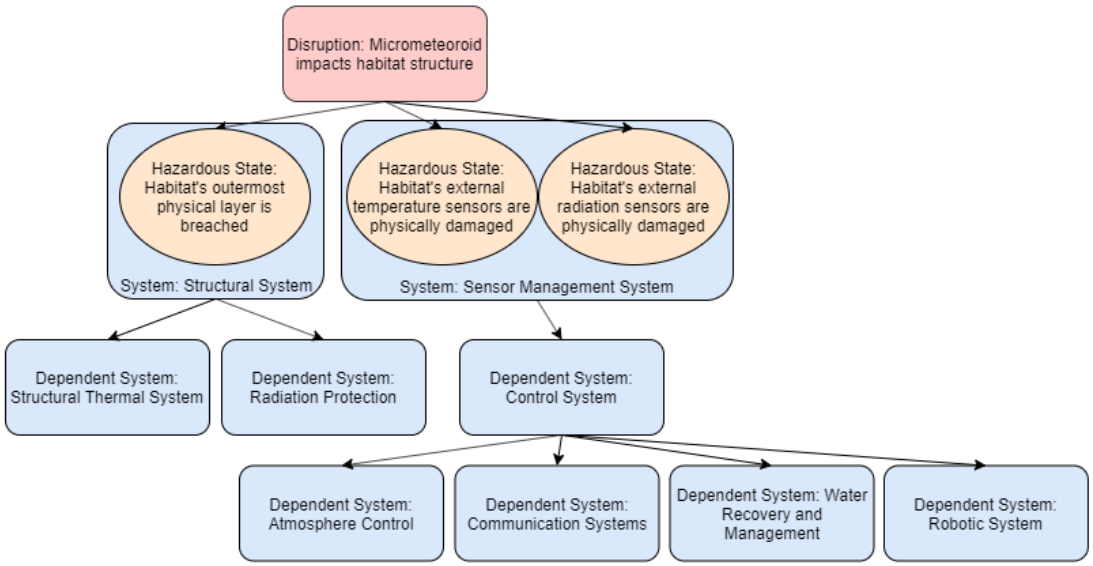

Figure 3.3: Disruption propagation through the habitat systems

Our identification of disruptions and hazardous states quickly resulted in a large amount of data. We need a method of storing not only the data, but the relationships between disruptions and hazardous states, and hazardous states resulting in other hazardous states. To reflect these relationships for a large amount of data, we use Microsoft Access to store the data in tables that are organized in a way that allows us to link relationships between cells where appropriate. The data is formatted for export to Microsoft Excel, and then these Excel spreadsheets are loaded into Matlab where we use the data to create a directed network. This exercise was completed with the help of the two undergraduate aeronautical and astronautical engineering students mentioned previously, and we describe the format of this database and network in the next section.

3.2 The Database of Disruptions and Hazardous States and the Failure Network

The database is structured into eight distinct tables. These tables are structured in the format shown in Figure 3.4, where the dashed boxes represent the data contained in each table.

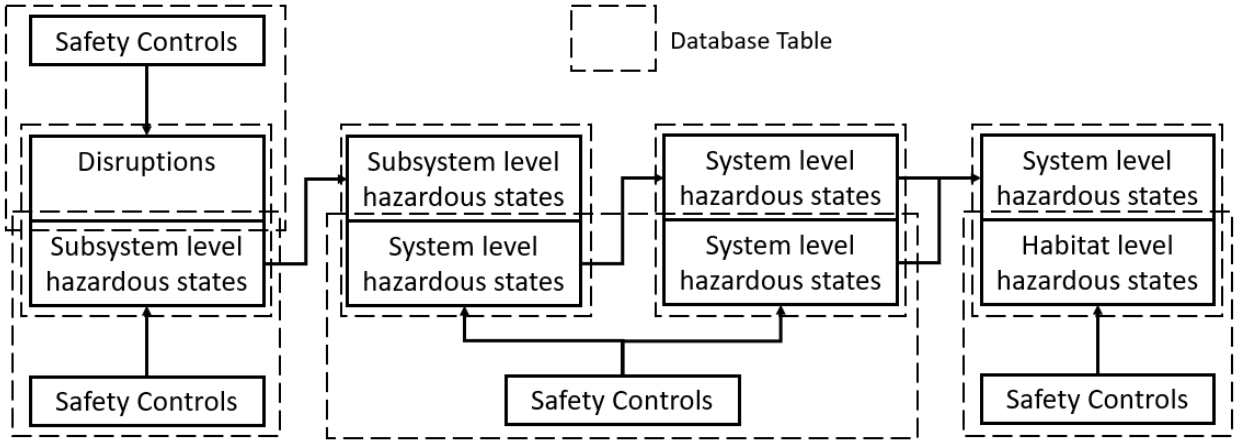

Figure 3.4: Database Structure

In this chapter, we focus on the four tables that do not include safety controls. These four tables are designed specifically to organize the disruptions and hazardous states. The tables are structured so that we follow a specific format when creating relationships. Hazardous states are organized in three levels: subsystem level, system level, and habitat level. Habitat level hazardous states include states that directly affect the living conditions in the habitat, such as *Hazardous chemicals present in habitat*, *Internal temperature above livable condition*, and *Habitat has no electrical power*. In Figure 3.3, we show how a disruption propagates through the habitat based on habitat system dependencies, and we reflect this in the database structure. First, the nominal state always transitions to a subsystem hazardous state through one of the 19 identified disruptions. Then, based on what system the affected subsystem is a part of, the subsystem hazardous state triggers transition to that system's hazardous state. Next, based on the system dependencies shown in Figure 3.2, that system hazardous state triggers transition to any other system hazardous state that has an identified dependency on the original system hazardous state. Also, we consider any habitat level hazardous states that may occur from any of the identified system level hazardous states and link those accordingly. We result in a layered database structure where the habitat transitions from nominal state to disruption, from disruption to any identified subsystem level hazardous state, subsystem level hazardous state to its associated system level hazardous state, from system level hazardous state to any other system level hazardous states based on system level dependencies, and finally from system level hazardous state to any identified habitat level hazardous states. To store these relationships, we create four separate tables. The first table links disruptions to their associated

subsystem level hazardous states. The second table links the subsystem level hazardous states to their associated system level hazardous states. The third table links the system level hazardous states to their associated dependent system hazardous states. The fourth table links the system level hazardous states to their associated habitat level hazardous states. Once these relationships are created in Microsoft Access, we export the tables to Excel and then we load the data into Matlab for ease of manipulation and visualization.

We define the failure network as the directed network that links the nominal state, disruptions, and associated hazardous states. The safety controls are not included in the failure network. The network is formatted in a layered orientation to illustrate the propagation of the habitat from the nominal state to the disruption, and through the three levels of hazardous states resulting in a habitat level hazardous state. Figure 3.5 shows the current failure network. The failure network illustrates the propagation from the nominal state (blue node), to the disruptions (magenta nodes), to the hazardous states (red nodes). To simplify the diagram, where two hazardous states are connected by the *No Action* trigger, we connect them directly.

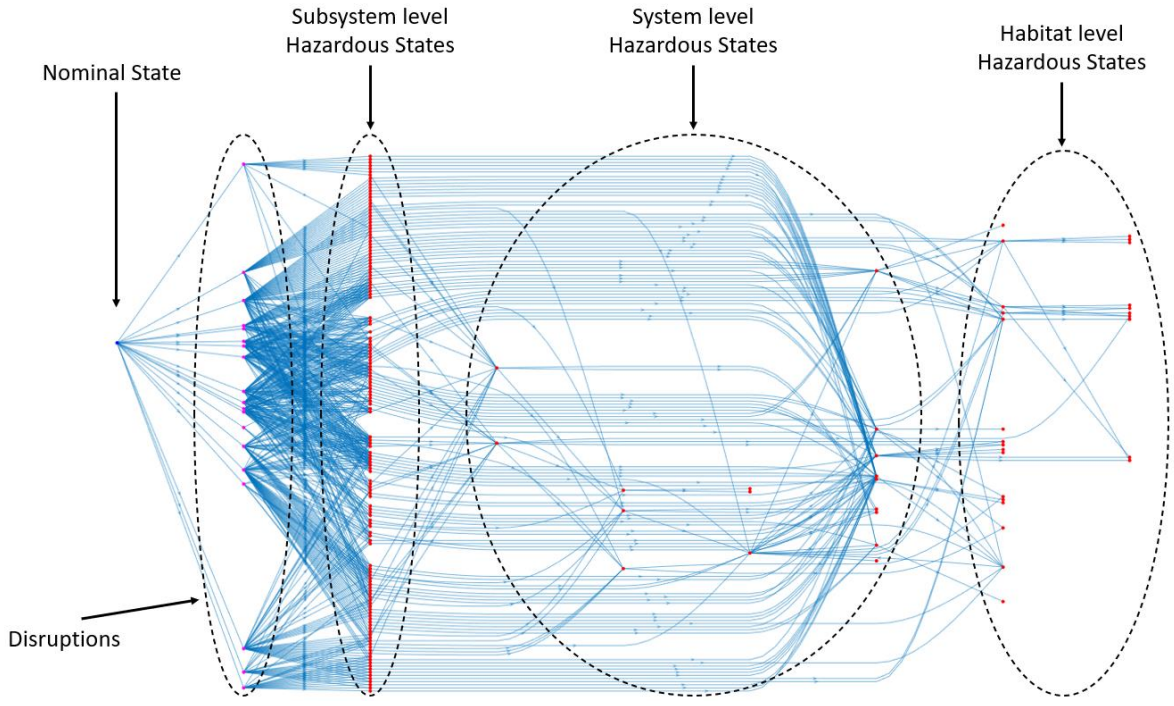

Figure 3.5: Failure Network with labeled hierarchical groupings

The failure network by itself is useful to evaluate certain network measures that we can use to investigate the relationships between the nodes. For example, which disruptions result in the most hazardous states? Which hazardous states have the most links to them? Which hazardous states are more central to the network, meaning if we controlled a certain hazardous state would that have any effect on other possible hazardous states? This evaluation is covered further in Chapter 4.

4. ASSESSING DISRUPTIONS AND HAZARDOUS STATES

In this chapter, we focus on hazard analysis and the second step in our control-theoretic approach to resilient design, assessing disruptions and hazardous states. There are many established ways to assess hazards in risk management, including a Preliminary Hazard Analysis (PHA), System Hazard Analysis (SHA), and Operations & Support Hazard Analysis (O&SHA). It is also common to use reliability techniques such as Failure Modes and Effects Analysis (FMEA) and Fault Tree Analysis (FTA) to investigate failures and combinations of failures that may result in hazardous effects to the system. We briefly cover the application of these techniques, and the inclusion of these techniques in our control-theoretic approach. We also discuss how we can use our failure network and associated network metrics to investigate the relationships between different disruptions and hazardous states, and also to investigate which disruptions and hazardous states have the most effect on other disruptions and hazardous states.

4.1 Traditional Hazard Assessment Techniques to Assess Disruptions and Hazardous States

While we use our network measures to investigate the relationships between our considered hazardous states and disruptions, we must also consider established methods of hazard assessment to view holistically how we prioritize controlling our hazardous states and disruptions. In this section we list a subset of these methods of hazard assessment and propose using them in combination with our failure network. The first method is a grouping of methods: preliminary, subsystem, and system hazard analysis. The primary objective of a hazard analysis is to identify all possible hazards, then categorize the hazards based on the severity of the consequences of the hazard (often catastrophic, critical, or marginal), and then evaluate the probability of the hazard occurring (Bahr, 2016). A preliminary hazard analysis is often the first hazard assessment technique conducted on the system, a subsystem hazard analysis identifies specific hazards and safety concerns for each major subsystem, and a system hazard analysis identifies specific hazards and safety concerns across subsystem boundaries and interfaces. A hazard analysis is very useful in ranking identified hazards, and in our case for ranking disruptions and hazardous states. For our use, we propose categorizing hazardous states and disruptions based on severity: catastrophic, critical, marginal, or negligible. Then, we categorize hazardous states and disruptions based on

probability: frequent, probable, occasional, remote, improbable, or eliminated. This is a common practice in industry, for example NASA defines *improbable* as a probability of less than 1% that a hazard will occur. Additionally, NASA defines "probable" as a probability of between 10% and 33% that a hazard will occur. These values are used in the Orion Crew Exploration Vehicle program (Perera, 2012). Finally, with each hazardous state and disruption having a severity and a probability, we can create a risk assessment matrix where we can group the hazardous states and disruptions accordingly. For example, a hazardous state with negligible severity and improbable probability does not prioritize control. Alternatively, a hazardous state with catastrophic severity and frequent probability requires immediate control. The combination of our failure network that categorizes disruptions and hazardous states into levels, along with a hazard analysis that can assess hazards at the preliminary, subsystem, and system level could provide a sufficient level of fidelity when considering which hazardous states and disruptions in our database provide the most risk.

Another hazard assessment technique that we consider for use in combination with our framework is the operations and support hazard analysis (O&SHA). We use O&SHA mainly to understand how hazards related to operations impact the system and to identify and evaluate these hazards. The O&SHA is conducted like the preliminary or system hazard analysis, but focuses on operations, concurrent task effects and limitations, human-machine-environment interfaces, and planned, unplanned, and hazardous operations. We identify hazardous states and disruptions to our operating habitat system, and we aim to design a system that is resilient in operation. Therefore, completing an O&SHA would be an excellent complement to our hazard identification and assessment techniques. Finally, we consider the use of a Failure Modes and Effects Analysis (FMEA). Used primarily as a reliability engineering tool, system safety engineers have used FMEA since the 1960s to identify failures in systems. FMEA can be used to investigate how a failure preceding a hazardous state can occur. Once other hazard analysis tools like a PHA or O&SHA have been used to identify hazardous states, an FMEA can be used on a case by case basis to focus on how particular failure modes might lead to and create a hazardous state (Bahr, 2016). Although an extremely powerful analytical tool, FMEAs are expensive to perform and can be laborious and tedious. We recommend that an FMEA be used to investigate the failures that may result in our identified hazardous states, or to identify new hazardous states based on identified component or subsystem failures.

In our framework, we make use of a network to list the disruptions and hazardous states that we have identified and to document the relationships between them. This is useful to understand the connectedness within our habitat system and the interfaces that are present. However, to fully understand and rank the disruptions and hazardous states, we recommend making use of established hazard analysis techniques like a PHA and the related SSHA and SHA, O&SHA, and FMEA. By doing this, we can assign levels of risk to our identified hazardous states and disruptions, possibly identify new hazardous states and disruptions, and use these levels of risk when we evaluate our safety controls and their effectiveness.

4.2 Application: Network Theory and the Failure Network

In this section we briefly introduce the concept of network theory and how we are applying several concepts from this field to our database structure. A network, also called a graph in much of the mathematical literature, is a set of nodes (also called vertices) with edges that connect between them (Newman, 2003). Networks have been used to study the World Wide Web, social networks, business relations, and were brought to mainstream attention through Milgram's Small World Theory which suggested that the human network is a small-world type network, where the average degree of separation between two humans is six (Newman, 2003). Recent network research has focused on the consideration statistical properties of large-scale graphs. This is mainly due to the availability of new technology that allows us to gather and analyze data on a scale far larger than previously possible. It is common to analyze networks that contain millions or even billions of vertices. This development of statistical methods for quantifying large networks lends tools that we can use to analyze our failure network. Also, because our failure network is on a much smaller scale (hundreds, rather than millions of nodes), we can use the power of the human eye. Using the human eye to analyze networks is an excellent way to understand their structure (Newman, 2003). Network theory provides methods that we can use to analyze centrality (which nodes are best connected to others or have the most influence) and connectivity (whether and how nodes are connected to one another in the network), in addition to using our visual judgment of the network.

A network is a set of nodes, connected by edges. Edges are defined as directed, or undirected. A directed edge runs only in one direction between two nodes, whereas an undirected edge runs in both directions. Directed edges are often indicated by arrows, as in our failure network. A graph

is directed if all its edges are directed. Our failure network, shown in Figure 3.4, is a directed network. We have assigned every edge in our failure network a direction, and no paths in the failure network are bidirectional. Thus, we will only use network measures that are specific for analyzing directed graphs for our failure network. When we analyze our failure network using network theory techniques, we omit safety controls because we are only interested in how the disruption and hazardous state nodes are connected and related in terms of network centrality.

The first network measure we consider is **node indegree**. The indegree of a node measures the number of incoming edges to that node. For example, in our failure network, omitting safety controls, the subsystem level hazardous state *Habitat thermal protection layer is physically damaged* has an indegree of 10. This indicates that 10 of the 19 identified disruptions result in that subsystem level hazardous state because currently the only nodes that are connected to subsystem level hazardous states are disruptions. We use node indegree to measure the amount of disruptions or hazardous states that result in a particular hazardous state.

The second network measure we consider is **node outdegree**. The outdegree of a node measures the number of outgoing edges from that node. For example, in our failure network the disruption *Micrometeoroids impact habitat* has an outdegree of 37. This indicates that we have identified 37 unique subsystem level hazardous states that could occur from the impact of a micrometeoroid on the habitat. In our network structure, node outdegree is especially useful for analyzing disruptions. By measuring a disruption's outdegree, we know exactly how many hazardous states result directly from that disruption. Thus, from a network perspective, the disruptions with the highest outdegree have a larger chance of creating hazardous states than the disruptions with the lowest outdegree. Thus, we consider outdegree to provide an indication in our analysis of which disruptions to consider when implementing safety controls. We can also use node outdegree to measure how many hazardous states result from a particular hazardous state, and in combination with node indegree is useful in analyzing the failure network relationships. A simple diagram that shows the application of indegree and outdegree to an example node *k* is shown in Figure 4.1.

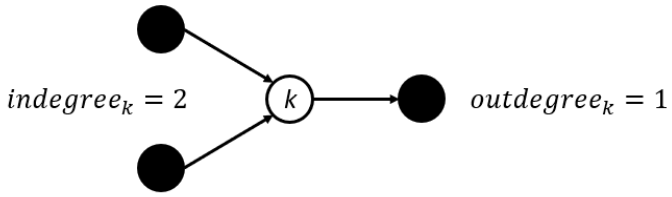

Figure 4.1: Indegree and Outdegree of Example Node *k*

Using these two network measures we can quantifiably measure the connectedness of the disruptions and hazardous states in our failure network. Also, because our failure network is relatively small, in the hundreds of nodes, compared to the scale of some networks being analyzed in current network research, often in the millions of nodes, we are still able to use the human eye test. Figures 4.2 and 4.3 show the failure network with the node size adjusted according to indegree and outdegree network measures.

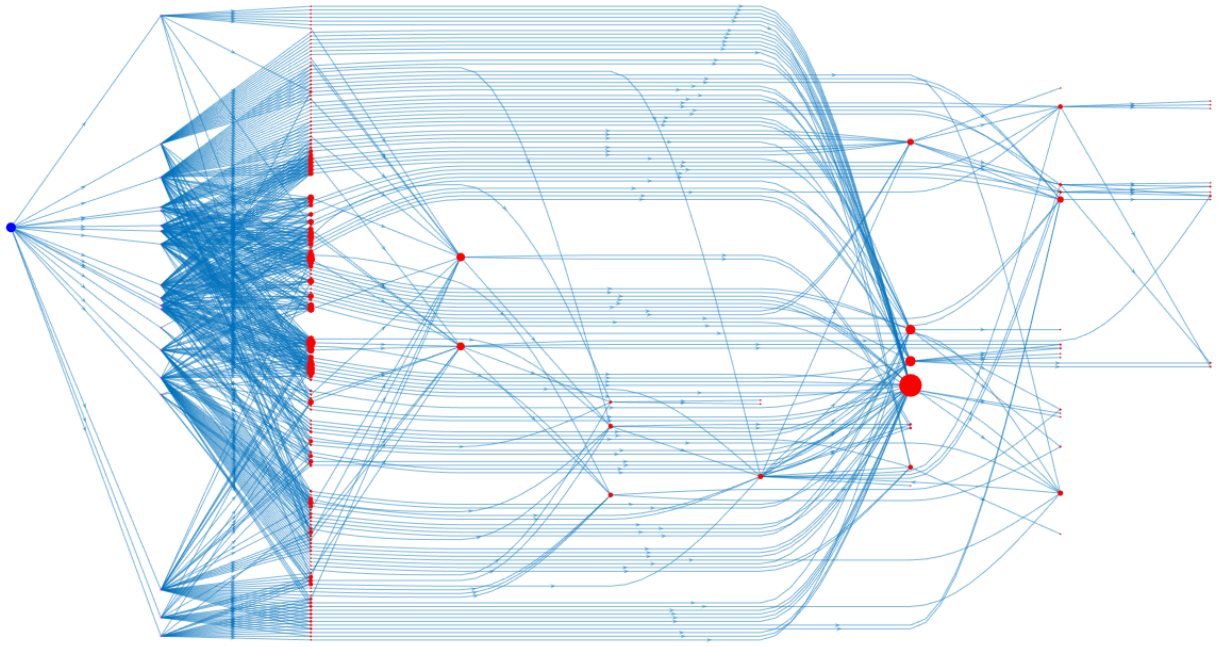

Figure 4.2: Failure Network, node size adjusted by node indegree

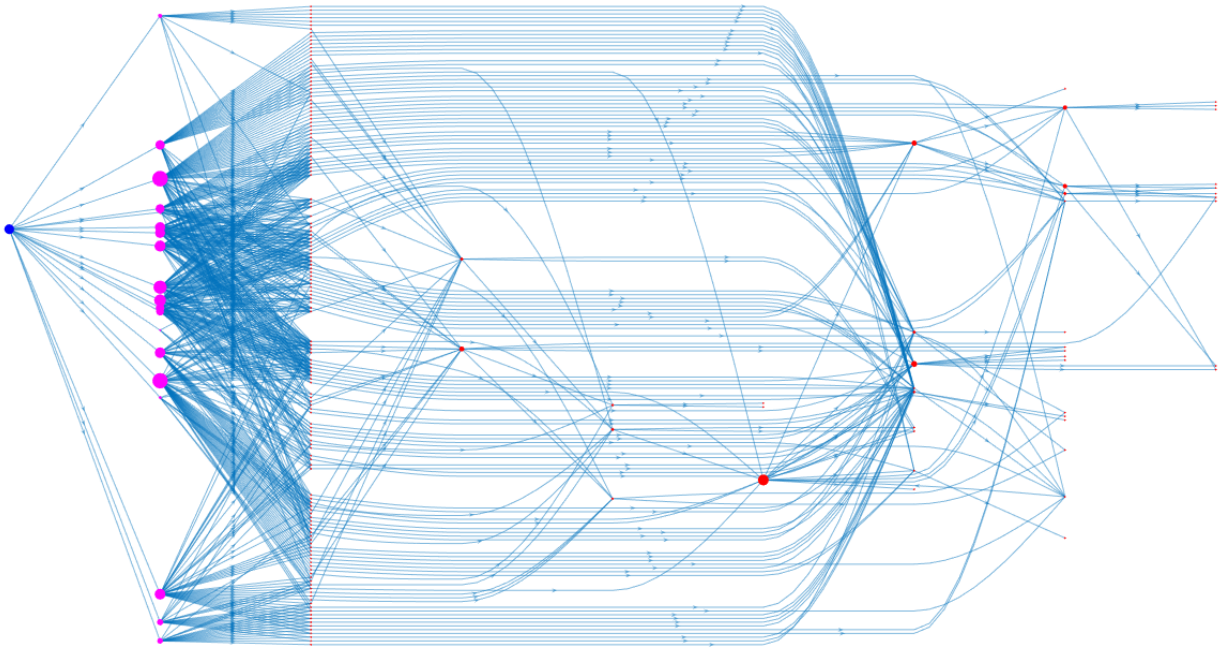

Figure 4.3: Failure Network, node size adjusted by node outdegree

By adjusting the node size based on the network measure that we want to consider, we can immediately see which disruptions and hazardous states to consider further. For example, in Figure 4.2, one system level hazardous state stands out as the largest node. This hazardous state, *Sensor system is not functioning properly*, has an indegree of 35. In other words, 35 hazardous states can lead to that hazardous state. In Figure 4.3, we see multiple disruptions that have high outdegrees, but we also note that one disruption, *Cold welding causes mechanical parts to fuse*, has the lowest outdegree. We should then, from a network perspective, consider that disruption with less urgency than other disruptions with larger outdegrees (all else being equal). Creating this network format for our database of disruptions and hazardous states allows us to visualize and analyze the relationships between the disruptions and hazardous states visually, or qualitatively, as well as with quantifiable values. We use these values to understand which disruptions and hazardous states may be the most critical to control.

To select a hazardous state or disruption to control, we use two factors: specific content, and network measures. For example, if a designer is only interested in the implications of radiation to the habitat, we would isolate the radiation disruption. We choose a disruption to control based on that node's specific content. However, if a designer wants to look holistically at the different

disruptions and hazardous states to determine which ones have the most connections, we use our network measures to assign numeric values to each disruption or hazardous state. For disruptions, we identify each node's outdegree, or how many edges originate from that node that result in hazardous states. The disruptions we focus on controlling are the disruptions with the highest outdegree. For hazardous states, we consider both indegree and outdegree to measure a hazardous state's network risk criticality. Currently, there exists in the network relationships between disruptions and subsystem level hazardous states, subsystem level hazardous states and system level hazardous states, and system level hazardous states and other system level hazardous states and habitat level hazardous states. We have ensured capability in the network to allow for relationships between hazardous states at the same level, and for relationships from a top level to a lower level. For example, we envision that a subsystem level hazardous state can and will cause another subsystem level hazardous state, and that a habitat level hazardous state could cause a subsystem level hazardous state. Although these relationships are not currently reflected in the network, we maintain the capability to input these relationships as our database of disruptions and hazardous states grows. We also can reflect accident states in the network, and these should be considered when assessing a hazardous state's criticality. Thus, we maintain that we can investigate a hazardous state's importance within the failure network by considering the amount of hazardous states or disruptions that result in that hazardous states and the amount of hazardous states resulting from that hazardous state. Equation 4.1 shows our network risk criticality measure, which considers that the risk criticality of a hazardous state in the network is the sum of the hazardous state's indegree and outdegree.

$$Network\ Risk\ Criticality_{HS} = indegree_{HS} + outdegree_{HS}$$

Equation 4.1: Network Risk Criticality

Equation 4.1 considers both incoming and outgoing relationships. Setting up the equation this way allows us to uniformly consider the risk criticality of a certain hazardous state as the database grows, and we input more relationships. For example, a subsystem level hazardous state may result in other subsystem level hazardous states, system level hazardous states, and habitat level hazardous states. That subsystem level hazardous state may also result from many disruptions. Using Equation 4.1, we reflect that that hazardous state, because of its large amount of

relationships, has a high network risk criticality. Additionally, a habitat level hazardous state that results from many system level hazardous states may result in many accident level hazardous states once these are added to the network, and in that case would have a large network risk criticality. For disruptions, we consider only outdegree. All disruptions have 1 incoming relationship from the nominal state, and a more representative measure of their criticality is to consider how many hazardous states result from that disruption (i.e. that disruption's outdegree).

4.3 Assessing Disruptions and Hazardous States with Network Theory and Traditional Hazard Assessment

We have introduced two network measures to prioritize the use of safety controls for disruptions and hazardous states in our failure network. These network measures are good indications of the relationships between the disruptions and hazardous states, and in practice we recommend they be used in combination with existing hazard analysis techniques like a Preliminary or System Hazard Analysis, Operations and Support Hazard Analysis (O&SHA), or a Failure Modes and Effects Analysis (see Section 4.1). We consider the assessment of disruptions and hazardous states at two levels. First, we consider the local importance of a disruption or hazardous state based on its combination of probability and consequences. This can be achieved through traditional hazard assessment techniques. Second, we use our failure network to consider the relationships between different disruptions and hazardous states to relate how many other hazardous states might result from a certain hazardous state or disruption. As the habitat design progresses and our database grows, we may consider the local importance, using traditional hazard assessment techniques, of the cascading disruptions and hazardous states identified using the network measures. In this section, we use our network measures to assess disruptions and hazardous states and recommend when to use traditional hazard analysis techniques to prioritize the control of the disruptions and hazardous states.

For the disruptions in our failure network, we consider outdegree, or the number of subsystem level hazardous states resulting from a disruption, to be the distinguishing factor for the disruptions that we consider to be the most critical. That is, the more subsystem level hazardous states resulting from a disruption, the more critical it is to the habitat failure network. Table 4.1 shows the 19 considered disruptions and their associated outdegree in the failure network.

Table 4.1: Considered Disruptions and their Failure Network Outdegree

Disruption	**Outdegree**
Extreme low external temperature	54
Ionizing Radiation	53
Chemical Spill	45
Batteries explode	40
Micrometeoroid impacts habitat	37
Rapid rise in external temperature	36
Electrostatic discharge	35
High Pressure vessel rupture	34
High winds cause debris to impact the habitat	33
Rapid decrease in external temperature	31
Pressurized line break	29
Impact of ejecta	28
High winds cause dust to impact habitat	25
Extreme high external temperature	19
Non-ionizing radiation	17
Outgassing of materials	11
Seismic activity within/near habitat	10
Batteries overcharge	8
Cold welding causes mechanical parts to fuse	4

Using this data, we show that low temperatures, radiation, micrometeoroids, and operations disruptions like chemical spills and batteries exploding have significant implications for the number of identified subsystem level hazardous states in our failure network. Although the disruptions with fewer resulting subsystem level hazardous states should not be ignored, we note that disruptions like cold welding, outgassing, and non-ionizing radiation have less far reaching implications to the habitat than the disruptions with higher outdegrees. We use this network measure to prioritize disruptions to control.

For example, we assess the local impact of a disruption by estimating its probability and severity, and thus identifying its criticality to the habitat. There are many ways to assess local criticality that we discussed in Section 4.1, and one way is to consider a Preliminary Hazard Analysis (PHA). We can assign categories of severity, including catastrophic, critical, marginal, or negligible. We also consider categories of qualitative probability, such as frequent, probable, occasional, remote, improbable, and eliminated. In a PHA, the Hazard Risk Index (HRI) is used to indicate how the hazard can be addressed. For example, if a hazard is probable and has a high severity, it must be controlled immediately. A common practice is to set up a risk matrix that specifies the HRI for each hazard, and we encourage use of this practice to assess the local impact of each of our disruptions and hazardous states. In our space habitat example, ionizing radiation is a continuously

present disruption within the extraterrestrial environment and involves critical health concerns for the crew members. Thus, ionizing radiation should be considered for immediate control. Also, using our network measures, we show that ionizing radiation leads to 53 subsystem hazardous states, or about a third of all total subsystem hazardous states. Therefore, ionizing radiation can cause a large impact on the habitat system when we consider relationships between disruptions and hazardous states in isolation, and also if we consider the disruption in isolation.

For the subsystem level hazardous states, we use network risk criticality, or the sum of the number of disruptions that result in a specified hazardous state and the number of system level hazardous states that result from that subsystem level hazardous state, to be the distinguishing factor for the subsystem level hazardous states that we consider to be the most critical. We have identified 143 subsystem level hazardous states, and so Table 4.2 shows the five subsystem level hazardous states with the highest network risk criticality and the five subsystem level hazardous states with the lowest network risk criticality.

Table 4.2: Excerpt of Subsystem Level Hazardous States and their Network Risk Criticality

Subsystem Level Hazardous State	**Network Risk Criticality**
Structural seals are physically damaged	15
Habitat's outermost physical layer is damaged	13
External temperature sensors are physically damaged	12
External radiation sensors are physically damaged	12
Electrical circuit is physically damaged	12
⋮	⋮
Habitat thermostat not functional	2
Habitat interior valves not functional	2
Habitat radiation protection performance decreased	2
Solar arrays below nominal temperature	2
Solar arrays are covered by frost	2

Using this data, we show that the more critical hazardous states involve physical damage to the exterior of the habitat and to the habitat sensors. Currently, a subsystem level hazardous state's network risk criticality is dominated by its indegree, or the amount of disruptions that result in that subsystem level hazardous state. As more relationships are added, subsystem level hazardous states will result in more hazardous states, increasing the outdegree and subsequently increasing the network risk criticality.

For system level hazardous states, we also consider network risk criticality to prioritize these hazardous states. Table 4.3 shows the system level hazardous states and their associated network risk criticality.

Table 4.3: System Level Hazardous States and their Failure Network Betweenness Centrality

System Level Hazardous State	**Network Risk Criticality**
Sensor system not functioning properly	36
Mechanical system not functioning properly	21
Power distribution system not functioning properly	19
Thermal management system not functioning properly	17
Structural system not functioning properly	16
Power generation system not functioning properly	15
Water recovery and management system not functioning properly	14
Air quality system not functioning properly	10
Food management system not functioning properly	10
Power storage system not functioning properly	9
Communication system not functioning properly	9
Oxygen generation system not functioning properly	8
Control system not functioning properly	8
Structural-Thermal system not functioning properly	8
Atmospheric CO_2 removal system not functioning properly	7
Lighting system not functioning properly	7
Radiation protection system not functioning properly	7

Using this data, we show that the systems with the most interfaces, when degraded have significant impact on our failure network. For example, the power distribution system provides power to the rest of the systems in the habitat. If the power distribution system is degraded, this would have a significant and cascading effect on the rest of the systems in the habitat. Also, the sensor system has a large risk based criticality because a failure of the habitat sensors, how the crew monitors different habitat health states and critical values like temperature, pressure, and radiation, would have a significant effect on how the rest of the systems respond to hazardous states. Another example is that the radiation system and structural-thermal system have low network risk criticality. This is because, although these systems are integral to the habitat design, their interfaces are highly localized. A failure of either of these systems would be significant to the other system, but it would not have much of an impact on the rest of the considered systems. As more relationships are added to the database, especially within levels, the network risk criticality will adjust to reflect these relationships. We reiterate that using traditional hazard assessment techniques to measure local importance based on probability and severity will allow a holistic assessment of the relationship between disruptions and hazardous states and their relative hazard importance.

For habitat level hazardous states, we also consider network risk criticality. In the current network, habitat level hazardous states occur in transition from system level hazardous states. However, as our network grows, we can input relationships between disruptions and habitat level hazardous states, and relationships between habitat level hazardous states and possible accident states. Table 4.4 shows the habitat level hazardous states and their associated network risk criticality.

Table 4.4: Habitat Level Hazardous States and their Network Risk Criticality

Habitat Level Hazardous State	**Network Risk Criticality**
Habitat has reduced electrical power	3
Habitat has no electrical power	3
Internal habitat pressure is too low	3
Airborne dust is present within the habitat	2
Internal habitat pressure is rapidly decreasing	2
Internal habitat pressure is slowly decreasing	2
Oxygen concentration in habitat is too low	2
Internal temperature is above livable condition	2
Internal temperature is below livable condition	2
Internal habitat pressure is too high	2
Crew is trapped inside of habitat	1
Crew does not have edible food	1
Crew cannot communicate with Earth	1
Crew cannot communicate with each other	1
Crew is trapped outside of the habitat	1
Open flame in habitat	1
Oxygen concentration in habitat is too high	1
Crew visibility is impaired	1
Hazardous chemicals are present in the habitat	1
Carbon dioxide level in the habitat is too high	1
Habitat humidity is too high	1
Habitat humidity is too low	1
Electric charge buildup is present on the exterior of the habitat	1
Internal radiation level is above a livable condition	1

We show in Table 4.4 that the network risk criticalities of the habitat level hazardous states are low compared to other hazardous states and are very similar to each other in value. This is because of how the failure network is currently structured, and that habitat level hazardous states are the endpoint of the failure network and can only occur via habitat level hazardous states. Our network risk criticality measure and the network format allow for the addition of hazardous states and disruptions, and for the addition of connections between hazardous states of the same level. As our database grows, the network risk criticality of habitat level hazardous states will increase as more connections to other disruptions and hazardous states will be documented, as well as connections to habitat level hazardous states.

In this chapter, we focused on hazard analysis and the second step in our control-theoretic approach to resilient design, assessing disruptions and hazardous states. We briefly covered the application of traditional hazard assessment techniques, and the inclusion of these techniques in our control-theoretic approach. We also discussed how we can use our failure network and associated network metrics to investigate the relationships between different disruptions and hazardous states, and also investigated which disruptions and hazardous states have the most effect on other disruptions and hazardous states. We proposed a two-pronged approach of the assessment of disruptions and hazardous states. First, we propose considering the local importance of a disruption or hazardous state based on its combination of probability and consequences. Second, we proposed using our failure network to consider network risk criticality of different disruptions and hazardous states to relate how many other hazardous states might result from a certain hazardous state or disruption. In this way, we can consider hazardous states that may have critical local consequence but have relatively few implications in the network format, and we can consider hazardous states that may not have critical local consequence but have a large set of relationships in the network format. Lastly, hazardous states that have critical local consequence and a large importance in the network can be prioritized for control.

5. USING SAFETY CONTROLS TO MITIGATE DISRUPTIONS AND HAZARDOUS STATES

In this chapter, we discuss how we identify and use safety controls to mitigate disruptions and hazardous states. This chapter covers step 3 in our control-theoretic approach to resilient design and covers steps 3a and 3b in our process for identifying safety controls (see Figure 3.1). These processes fall under the larger scope of *Hazard Controls*, a part of the system safety engineering process, shown in Figure 2.3.

To develop safety controls for the hazardous states and disruptions, we use our knowledge in systems engineering, system safety, and of past accidents and incidents to develop safety controls that are designed to address these hazardous states and disruptions. As safety controls are identified throughout the design process, we identify the underlying principle of each safety control and generate a corresponding generic safety control. Generic controls specify safety controls based on their method or principle of control. We use these generic safety controls to develop more safety controls for different kinds of disruptions and hazardous states, identifying and using principles from system safety engineering to categorize controls and expand their applicability. We term the resulting set of potential safety controls the *safety control option space*.

5.1 Application: The Safety Control Option Space and Generic Safety Controls for a Martian Habitat

We continue our example from Section 3.1, where we apply our process described in Figure 3.1 to an example Martian habitat. We have developed a database of disruptions and hazardous states, and in this section, we develop safety controls to address these disruptions and hazardous states.

Step 3a and 3b: Develop safety controls and generic safety controls

Table 5.1 shows how we use the disruptions and hazardous states from Figure 2.1 to identify safety controls.

Table 5.1: Identification of Safety Controls for Example Disruptions, Hazardous States, and Triggers

Disruption, Hazardous State, or Trigger	Safety Control
Disruption: Micrometeoroid breaches habitat structure	Habitat structural protection strong enough to withstand impact
Disruption: Dust storm impacts habitat	Ability to remove dust contaminants with humans or robot repair agents
Hazardous state: Weakened habitat thermal protection	Ability to increase heat output to meet temperature demand
Hazardous state: Power unit damaged, degraded functionality	Ability to use backup power source

The disruption may also propagate through the habitat, creating additional hazardous states and triggers to those states. These triggers and hazardous states, in turn, may be addressed with safety controls. The disruptions and their propagation quickly result in a large space of states, triggers, and potential safety controls. In the example, the initial list of 19 disruptions result in 186 interconnected hazardous states. Although the final set of selected safety controls will not necessarily directly address each hazardous state or disruption, the total number of states and disruptions (205) provides a reasonable initial estimate of the potential size of the safety control option space. The development of this safety control option space is discussed next.

5.1.1 The Safety Control Option Space for an Example Martian Habitat

Described in Chapter 2, a safety control is any part of the system design or operation that maintains the system in a nominal state, prevents the system from propagating to a hazardous state, or restores the system from a hazardous or accident state to a nominal state. Safety controls may be active, or safety controls that respond to a disruptive event (e.g., by performing a repair), or they may be passive, or built into the design (e.g., thicker protections, operational and physical redundancies). We use generic safety controls, referenced in Figure 3.1, and described in the next section, to categorize controls and expand their applicability.

Some example safety controls are identified in Figure 2.1 and Table 5.1, and we show in Figure 2.1 that we can develop safety controls to address disruptions directly, or address the hazardous state resulting from the disruption. We also show that we can develop multiple safety controls for a single disruption or a single hazardous state. The disruption may also propagate through the habitat, creating additional hazardous states as illustrated in Figure 3.3. These hazardous states, in turn, may be addressed with safety controls. We next show an illustrative example of how we use

our methodology to populate the safety control option space. This example is for the purpose of illustrating how we populate the safety control option space.

In Figure 5.1, we take an excerpt of the database under development and consider three disruptions that transition the habitat to the same hazardous state.

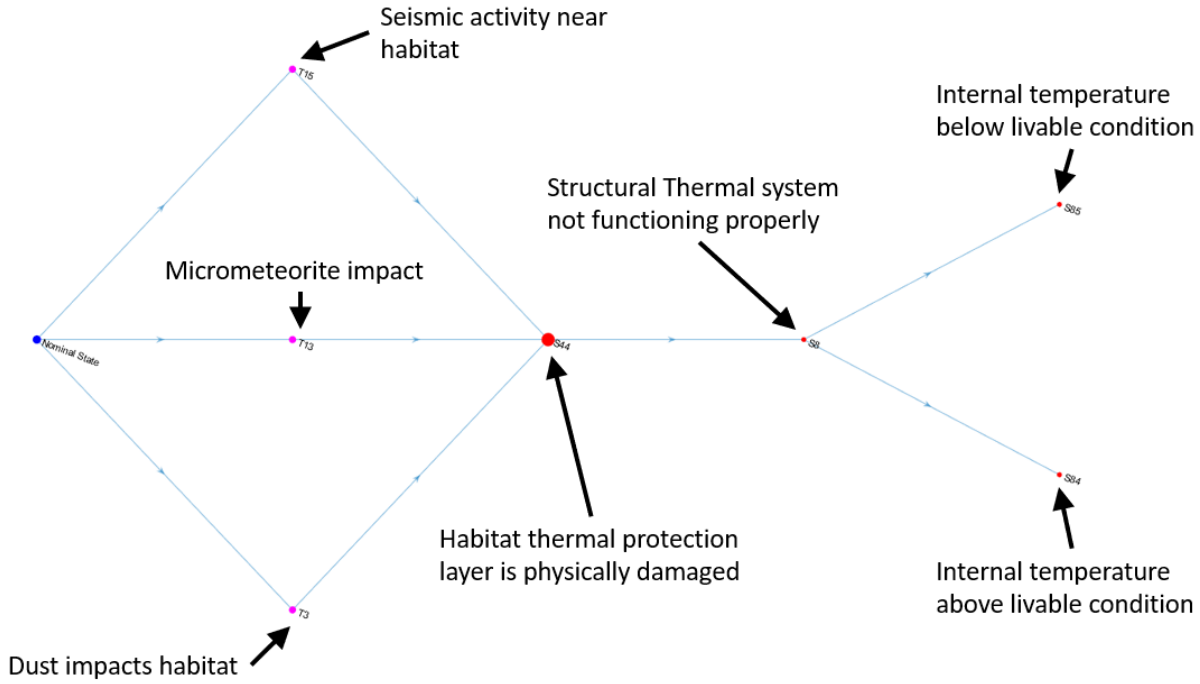

Figure 5.1: Excerpt of the Failure Network Considering Three Disruptions

The three disruptions shown Figure 5.1 are *Seismic activity near habitat*, *Micrometeorite impact*, and *Dust impacts habitat*. Considered individually, each one of these disruptions results in several hazardous states. For example, *Seismic activity near habitat* results in 10 hazardous states, *Micrometeorite impact* results in 37 hazardous states, and *Dust impacts habitat* results in 25 hazardous states. We note that there is some overlap, in that some of the considered disruptions directly result in the same hazardous states. Also, hazardous states that result from disruptions may cause additional hazardous states based on their system dependencies. In this example, we consider a small excerpt of the database and isolate one hazardous state that results from three distinct disruptions. This hazardous state, *Habitat thermal protection layer is physical damaged*, is a subsystem hazardous state that considers a subsystem (thermal protection layer) that is part of the

structural thermal system. Thus, the system hazardous state *Structural Thermal system not functioning properly* can be directly linked to the subsystem hazardous state *Habitat thermal protection layer is physically damaged*. Tracking system dependencies, we identify two additional hazardous states at the habitat level that may result from the system hazardous state *Structural Thermal system not functioning properly*. These habitat level hazardous states are *Internal temperature above livable condition* and *Internal temperature below livable condition*. Considering Figure 5.1, there are therefore several points in the network where intervention can prevent the eventual occurrence of either of the habitat level hazardous states. Table 5.2 shows examples of safety controls we identify along this path. We identify safety controls for each of the identified disruptions and hazardous states.

Table 5.2: Disruptions, Hazardous States, and their Safety Controls for Example Case Study

Disruption ID	**Disruption Description**	**Disruption Level**
T15	Seismic activity near habitat	N/A

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
T15-SC1	Build the habitat partially underground to avoid hinges and bearings of the habitat to the ground becoming loose	REMOVE COMPONENT FROM SOURCE	N/A
T15-SC2	Build the habitat fully underground to mitigate seismic activity impact	REMOVE COMPONENT FROM SOURCE	N/A
T15-SC3	Blend the habitat outer shell with Martian regolith to avoid discontinuities in structure that could become unhinged during seismic activity	REDUCE COMPONENT LOAD	N/A
T15-SC4	Habitat outer shell is strong enough to withstand seismic activity	COMPONENT WITHSTANDS SOURCE	N/A
T15-SC5	Use materials with high damping coefficients to mitigate seismic vibrations	COMPONENT CORRECTS FOR SOURCE	N/A
T15-SC6	Implement additional damping components to mitigate seismic vibrations	REDUCE COMPONENT LOAD	N/A

Table 5.2: Disruptions, Hazardous States, and their Safety Controls for Example Case Study (Continued)

Disruption ID	**Disruption Description**	**Disruption Level**
T13	Micrometeorite impact	N/A

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
T13-SC1	Habitat outer shell is strong enough to withstand micrometeoroid impact	COMPONENT WITHSTANDS SOURCE	N/A
T13-SC2	Habitat outer shell is thick enough to withstand micrometeoroid impact	COMPONENT WITHSTANDS SOURCE	N/A
T13-SC3	Ability to detect micrometeoroid impact and locally reinforce impact location	COMPONENT CORRECTS FOR SOURCE	N/A
T13-SC4	Ability to erect temporary outer shell, e.g. a blast shield, that could withstand micrometeoroid impact	REDUNDANT COMPONENT SYSTEM	N/A
T13-SC5	Locate the habitat underground so as to avoid micrometeoroids altogether	REMOVE COMPONENT FROM SOURCE	N/A
T13-SC6	Ability to provide rooms or compartments with extra structural protection to take cover during micrometeoroid impact	EXTRA PROTECTION FOR HUMANS FROM SOURCE	N/A

Disruption ID	**Disruption Description**	**Disruption Level**
T3	Dust impacts habitat	N/A

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
T3-SC1	Implement habitat outer shell that is strong enough to withstand dust impact	COMPONENT WITHSTANDS SOURCE	N/A
T3-SC2	Implement habitat outer shell that is thick enough to absorb dust impact	COMPONENT WITHSTANDS SOURCE	N/A
T3-SC3	Ability to forecast high winds and erect reserve outer wind shields	COMPONENT CORRECTS FOR SOURCE	N/A
T3-SC4	Implement automated cleaning or vacuuming of habitat outer shell	REMOVE SOURCE FROM COMPONENT	N/A

Table 5.2: Disruptions, Hazardous States, and their Safety Controls for Example Case Study (Continued)

Hazardous State ID	**Hazardous State Description**	**Hazardous State Level**
S44	Habitat thermal protection layer is physically damaged	Subsystem

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
HS44-SC1	Ability for robots to repair thermal protection	ROBOT REPAIR COMPONENT	Yes
HS44-SC2	Ability for humans to repair thermal protection	HUMAN REPAIR COMPONENT	Yes
HS44-SC3	Ability to increase thermal output to compensate for lost heat protection	COMPONENT CORRECTS FOR SOURCE	No
HS44-SC4	Ability for thermal protection layer to provide sufficient insulation when physically damaged	COMPONENT ROBUSTNESS	Yes
HS44-SC5	Ability for crew to wear thermal protection while thermal protection layer is damaged	EXTRA PROTECTION FOR HUMANS FROM SOURCE	No
HS44-SC6	Ability to move humans underground to protect from increased heat loss	EVACUATE CREW	No

Hazardous State ID	**Hazardous State Description**	**Hazardous State Level**
S8	Structural Thermal system not functioning properly	System

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
HS8-SC1	Ability for crew to move into "safe area" of habitat until breached area is repaired	EVACUATE CREW	Yes
HS8-SC2	Ability for crew to wear suits or thermal blankets	EXTRA PROTECTION FOR HUMANS FROM SOURCE	No
HS8-SC3	Available portable heaters	REDUNDANT COMPONENT FUNCTION	No
HS8-SC4	Ability for crew to repair breach	HUMAN REPAIR COMPONENT	Yes
HS8-SC5	Ability for robot agents to repair breach	ROBOT REPAIR COMPONENT	Yes
HS8-SC6	Ability to apply extra thermal coatings or MLI sheets to damaged area	COMPONENT CORRECTS FOR SOURCE	Yes

Table 5.2: Disruptions, Hazardous States, and their Safety Controls for Example Case Study (Continued)

Hazardous State ID	**Hazardous State Description**	**Hazardous State Level**
S85	Internal temperature below livable condition	Habitat

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
HS85-SC1	Availability of portable heaters	REDUNDANT COMPONENT FUNCTION	No
HS85-SC2	Ability to route heat to certain parts of the habitat	COMPONENT CORRECTS FOR SOURCE	No
HS85-SC3	Ability for crew to wear suits or thermal blankets	EXTRA PROTECTION FOR HUMANS FROM SOURCE	No

Hazardous State ID	**Hazardous State Description**	**Hazardous State Level**
S84	Internal temperature above livable condition	Habitat

Safety Control ID	**Safety Control Description**	**Generic Safety Control**	**Return to Nominal State?**
HS84-SC1	Ability to use colder exterior Martian temperature to cool habitat	COMPONENT CORRECTS FOR SOURCE	Yes
HS84-SC2	Availability of portable fans	REDUNDANT COMPONENT FUNCTION	No
HS84-SC3	Ability to route air conditioning to certain parts of the habitat	COMPONENT CORRECTS FOR SOURCE	No
HS84-SC4	Ability to shut off non-essential heat producing electrical systems	REDUCE COMPONENT LOAD	Yes
HS84-SC5	Availability of fans built into the habitat	REDUNDANT COMPONENT FUNCTION	Yes

Once we have identified the safety control option space for this subset of disruptions and hazardous states, we apply these safety controls to the relevant excerpt of the network shown in Figure 5.1. The resulting expanded excerpt of the network is shown in Figure 5.2.

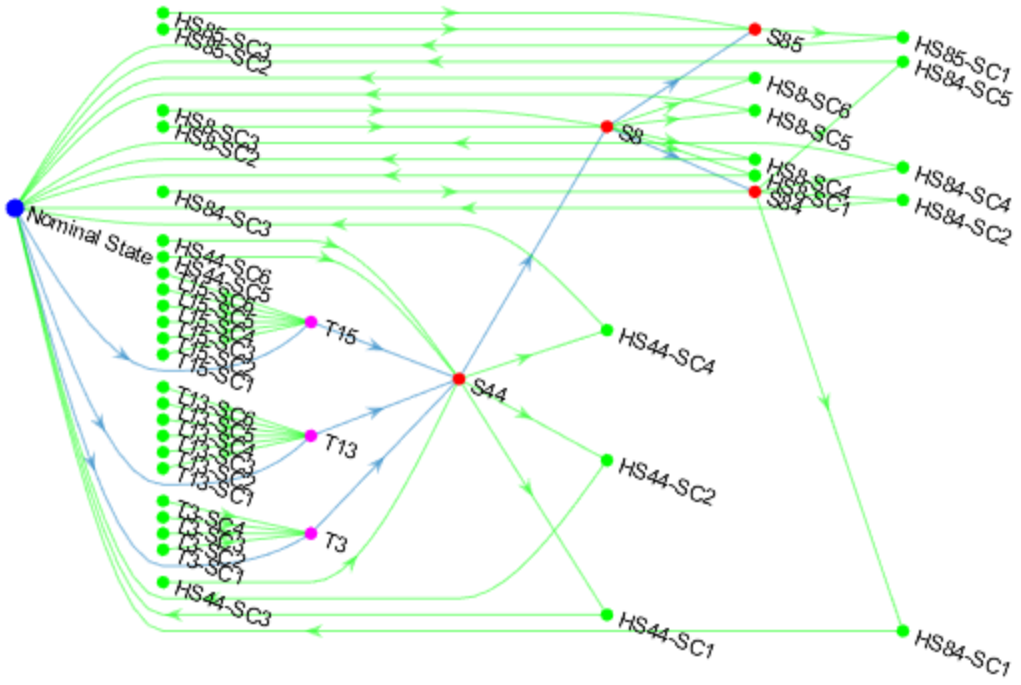

Figure 5.2: Safety Control Option Space Applied to an Excerpt of the Failure Network Considering Three Disruptions

We design safety controls for disruptions and at all levels of hazardous states. At the subsystem level, we aim to withstand a specific disruption, at the system level we ensure the habitat systems can prepare for a threat in advance and adapt during an encounter, and at the habitat level we ensure the safety of the crew while taking measures to mitigate the threat. Unlike Event Sequence Diagrams which track failures of basic components, safety controls facilitate creating layers of defense against disruptions at each level of the habitat.

5.1.2 Generic Safety Controls for an Example Martian Habitat

To develop the larger safety control option space to a current list of 776 safety controls, we repeat the process covered in the previous example for the expanded list of 19 disruptions and 186 interconnected hazardous states. We use the generic safety controls for two purposes. First, we use the underlying principle of control of the generic safety control to identify more safety controls with various applicability to different disruptions and hazardous states. Second, we use the generic safety controls as well as the safety controls to identify control flaws in step 4 of Figure 3.1 that we use to factor towards the control effectiveness metric, covered in Chapter 6. As we identify

safety controls, we use generic safety controls to identify more safety controls for different kinds of disruptions and hazardous states, identifying and using principles from system safety engineering to categorize controls and expand their applicability. In the previous example, many of the safety controls are specific implementations of the generic safety control *COMPONENT CORRECTS FOR SOURCE*. This generic safety control requires that a component or system can adapt or slightly change its functions in the face of a source of a disturbance. Table 5.3 describes all the current generic safety controls we have identified, along with an example of a specific safety control implementation of generic safety control.

For a consistent lexicon, we have defined terms that are used across several generic safety controls. **ROBOT** refers to the robot agent responsible for repairs, maintenance, inspections, and autonomous tasks that can be carried out without human intervention. **HUMAN** refers to the human agent, or crew member, responsible for maintaining, inspecting, and repairing the habitat when autonomous action is not sufficient or not possible. **COMPONENT** can refer to either the component, subsystem, or system in question that is being disrupted or is in a hazardous state. **SOURCE** refers to the source of the disruptive event or the hazardous state. For example, a dust storm is a source of a disruption, as is a micrometeorite impact.

Table 5.3: Generic Safety Controls

Generic Safety Control	**Generic Safety Control Description**	**Safety Control that applies Generic Safety Control Principle**
ROBOT REPAIRS COMPONENT	The robot agent carries out a repair on a component or system that is damaged or not functioning to full capacity	*Ability for robot agents to reinforce radiation protection*
HUMAN REPAIRS COMPONENT	The human agent carries out a repair on a component or system that is damaged or not functioning to full capacity	*Ability for human agents to repair adhesives*
REMOVE COMPONENT FROM SOURCE	In the presence or anticipation of a source, the component or system is removed or shielded from the source	*Ability to cover solar arrays in anticipation of dust storms*
REMOVE SOURCE FROM COMPONENT	The source of the disruption is removed	*Ability for solar arrays to vacuum or move dust*
COMPONENT WITHSTANDS SOURCE	The component or system can function at a necessary level in the presence of a source	*Ability for structure to withstand buildup of space dust*
COMPONENT CORRECTS FOR SOURCE	The component or system adapts its functions to protect against a source	*Ability to regulate temperature of solar arrays using heat exchangers*
REDUNDANT COMPONENT FUNCTION	The habitat can achieve the function of the component or system affected by the source using a different method	*Ability to use reserve power storage capabilities*
REDUNDANT COMPONENT SYSTEM	The habitat has another component or system to use when the component or system affected by the source cannot be used	*Ability to have a backup digital circuit*
REDUCE COMPONENT LOAD	The component or system affected by the source is used less or at a lower capacity to ensure functionality	*Ability to cycle through multiple external sensors to regulate their temperature and workload*
EXTRA PROTECTION FOR HUMANS FROM SOURCE	The habitat provides protection for the human agent from the source of the disruption	*Ability for crew to wear oxygen masks*
ISOLATE COMPONENT	The component or system affected by the source is isolated to prevent further hazardous states	*Ability to sequester breach from the rest of the habitat*
REPLACE COMPONENT	The component or system affected by the source is replaced	*Ability to replace exterior valves*
COMPONENT ROBUSTNESS	The component or system affected by the source is able to function in the presence or after being affected by a source	*Ability for sensors to work when above or below nominal temperature*
EVACUATE CREW	The human agents evacuate either to a part of the habitat that is not affected by the source, or leave the habitat entirely	*Ability to move crew to a part of the habitat that has not been breached*
RESUPPLY	The component, system, or resource produced or used by the component or system is resupplied from Earth	*Ability to shuttle water supplies*
HUMAN VERIFIES SOFTWARE	The human agent verifies and confirms a process done autonomously in the habitat	*Ability for humans to clearly understand and verify software results*
COMPONENT DECENTRALIZES FUNCTION	A component or system can work independently to achieve a function that is done by a centralized system	*Ability for systems to implement warnings independently of a warning system*

This list is not a complete list of the generic safety controls, and we expect to add to this list as we continue to identify disruptions, hazardous states, and safety controls. The following example shows how we use generic safety controls to identify more safety controls and expand our database, using two state and trigger models. Figure 5.3 shows a state and trigger model considering a disruption *Habitat cooling line leaks ammonia* taken from an incident on the International Space Station in 2013 (Dunbar, 2013).

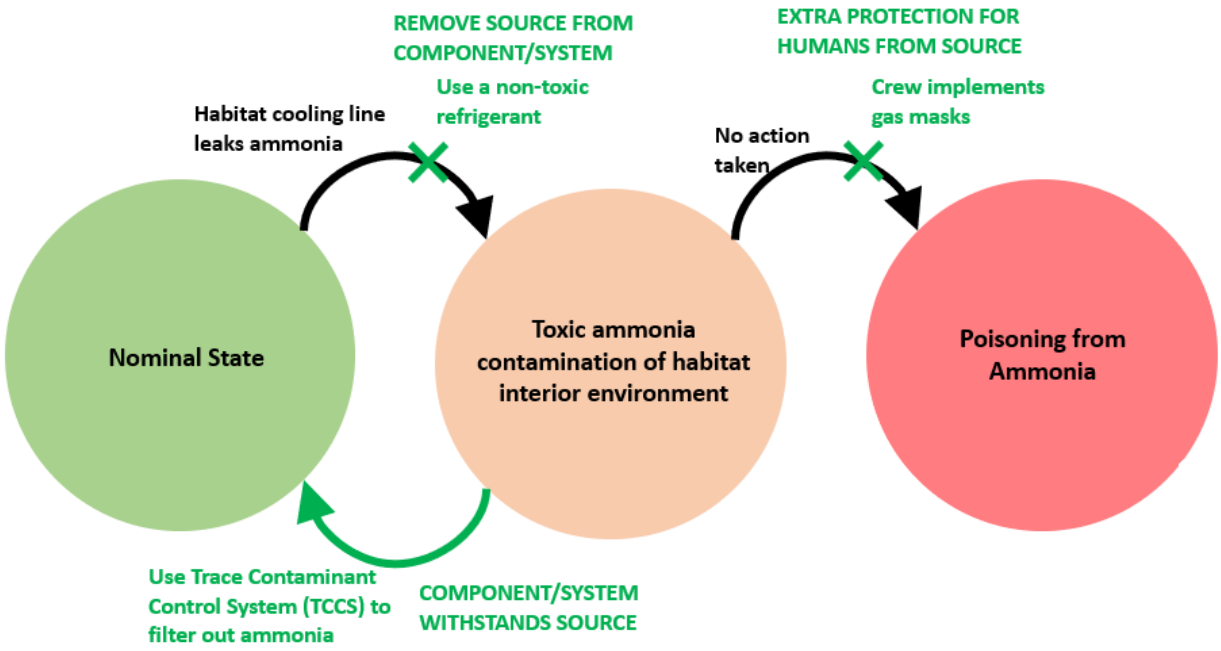

Figure 5.3: State and Trigger Model with Generic Safety Controls

In Figure 5.3, the habitat transitions from the nominal state to the hazardous state *Toxic ammonia contamination of habitat interior environment* through the disruption *Habitat cooling line leaks ammonia*. If no action is taken, the hazardous state transitions to an accident state *Poisoning from ammonia*. To prevent the transition from the nominal state to the hazardous state, we implement the safety control *Use a non-toxic refrigerant*. Because the source of the disruption is the leaking of a toxic refrigerant, ammonia, using a non-toxic refrigerant will remove the source of the disruption and prevent the hazardous state from occurring. Thus, this safety control identifies with the generic safety control *REMOVE SOURCE FROM COMPONENT/SYSTEM*. Further, to prevent the accident state, we implement the safety control *Crew implements gas masks*. This safety control provides protection for the human agents from the ammonia, so we identify it with the *EXTRA*

PROTECTION FOR HUMANS FROM SOURCE generic safety control. The final safety control we consider is to *Use Trace Contaminant Control System (TCCS) to filter out ammonia*. This is an example of a safety control where the habitat withstands the addition of a source of a disruption, so we identify this safety control with the *COMPONENT/SYSTEM WITHSTANDS SOURCE* generic safety control. Next, we use a different example of a disruption and hazardous state to show how the generic safety controls help create new safety controls, shown in Figure 5.4.

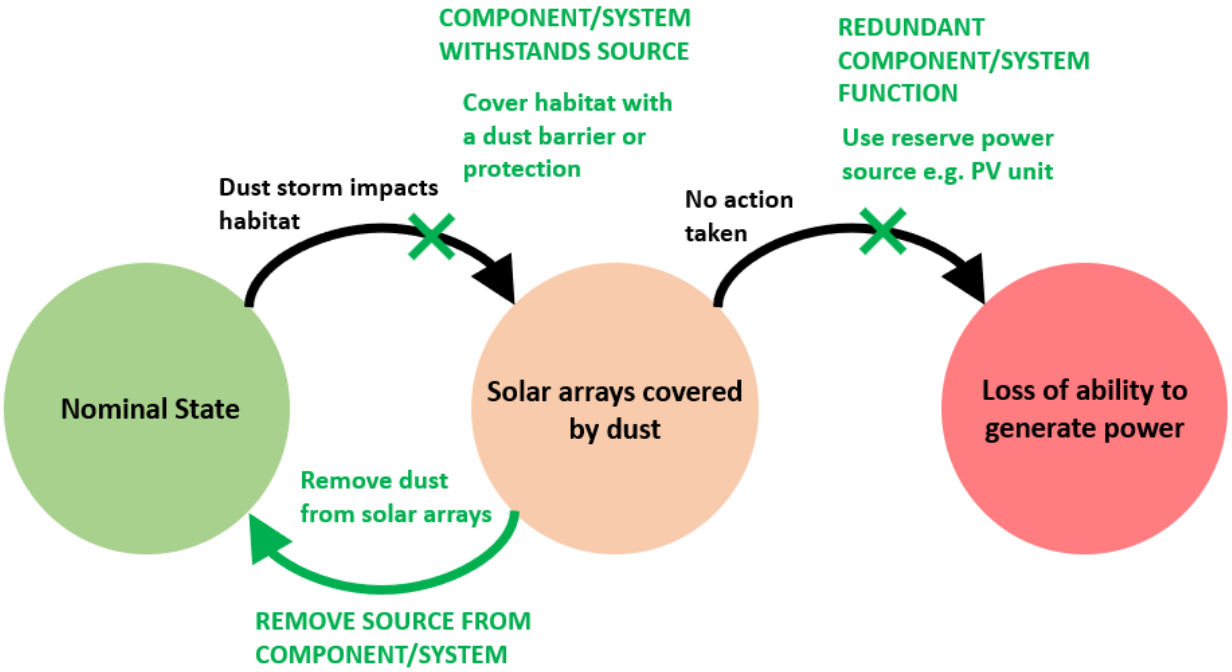

Figure 5.4: State and Trigger Model with Generic Safety Controls

In this example, we consider the habitat transitioning from the nominal state to the hazardous state *Solar arrays covered by dust* through the disruption *Dust storm impacts habitat*. Further, if no action is taken, the habitat transitions to the accident state *Loss of ability to generate power*. We use our generic safety controls from the previous example in Figure 5.3 and consider their applicability for this example in Figure 5.4. We identify that both *COMPONENT/SYSTEM WITHSTANDS SOURCE* and *REMOVE SOURCE FROM COMPONENT/SYSTEM* are applicable. To prevent transition to the hazardous state, we consider how we can implement *COMPONENT/SYSTEM WITHSTANDS SOURCE* and create the safety control *Cover habitat with a dust barrier or protection*. Similarly, we return the habitat from the hazardous state to the nominal state by implementing the safety control *Remove dust from solar arrays*, which we

generate by considering how to apply the generic safety control *REMOVE SOURCE FROM COMPONENT/SYSTEM*. To prevent transition to the accident state, we apply the safety control *Use reserve power source (e.g. PV unit)*. This safety control is an example of achieving the function of the component or system affected by the source of the disruption using a different method, and thus we can group this safety control in the *REDUNDANT COMPONENT/SYSTEM FUNCTION* generic safety control.

After completing this example, we have generated six total safety controls for two separate disruption scenarios, as well as creating and applying four distinct generic safety controls. We repeat this process for our current list of disruptions and hazardous states. When we identify that a created generic safety control is applicable to a disruption or hazardous state we are considering, we create a safety control that applies the control principle of that generic safety control. If a created generic safety control is not applicable, we create a new safety control and a new generic safety control principle. This has assisted in the current creation of 776 safety controls and 17 generic safety controls for the 19 disruptions and 185 hazardous states we have identified. We store these safety controls and generic safety controls in the safety control option space, stored in the relational database used to create the network structure.

5.2 The Failure Network and Storing Safety Controls

In this chapter, we have so far demonstrated how we develop the safety control option space. For each disruption and for each level of the hazardous states, we develop safety controls and link the safety controls accordingly to the disruptions or hazardous states that they are designed to control. In this section, we give an example state and trigger model and show how it is applied in the network format. Then, we extrapolate that process and show the full failure network with the safety controls included. We consider the following detailed example of a micrometeorite impact to the habitat. Table 5.4 details the resulting hazardous states and accident states, as well as the safety controls that we design to address the disruption and each hazardous state. In our state and trigger model, we cannot implement safety controls that return the habitat from an accident state to a nominal state. However, we can use safety controls that prevent an accident state from happening, and we can reduce damage once an accident state is reached. For example, having reserve space suits available for an accident state of *unlivable pressure environment* would prevent loss of life

of crew members, but the accident state has still been reached. Figure 5.5 shows the example in a state and trigger model.

Table 5.4: Hazardous and Accident States considered for *Micrometeorite impacts habitat.* HS = Hazardous State, AS = Accident State

Disruption		**Safety Control**	
Micrometeorite impacts habitat		Habitat outer shell is strong enough to withstand micrometeorite impact	
State ID	**State Description**	**Safety Control that prevents propagation from that Hazardous State**	**Safety Control that returns to Nominal State**
HS1	Structural seals are physically damaged	Ability to reinforce part of the structure with regolith	Ability for robots to repair structural seals
HS2	Structural system not functioning properly	Ability to isolate the breached part of the habitat	Ability for habitat to erect extra layer of protection
HS3	Internal habitat pressure is too low	Ability for crew to quickly apply their pressurized suits	Ability for ECLSS system to manually control for pressure regulation
HS4	Internal habitat pressure is slowly decreasing	Ability for crew to quickly apply their pressurized suits	Ability for robot agents to repair breach
HS5	Internal habitat pressure is rapidly increasing	Ability for crew to evacuate to "safe area" of habitat	N/A
HS6	Structural Thermal system not functioning properly	Availability of portable heaters	Ability to apply extra thermal coatings or MLI sheets to damaged area
HS7	Radiation protection system not functioning properly	Availability of an underground room where crew can shelter from radiation	Ability to run water through the habitat shell to reduce radiation impact
HS8	Internal temperature below livable condition	Ability to route heat to certain parts of the habitat	N/A
HS9	Internal temperature above livable condition	Ability to route cold air to certain parts of the habitat	Ability to use colder exterior temperature to cool habitat
HS10	Internal radiation level above livable condition	Ability to apply extra radiation protective suits	N/A
HS11	Electric charge buildup is present on exterior of habitat	Ability to keep sensitive electronics protected in case of electrostatic discharge	Ability to safely discharge electric buildup on habitat
AS1	Unlivable pressure environment	N/A	N/A
AS2	Unlivable temperature environment	N/A	N/A
AS3	Radiation poisoning	N/A	N/A
AS4	Loss of power due to electric discharge	N/A	N/A

We use the information in Table 5.4 to create a state and trigger model of the example, shown in Figure 5.5.

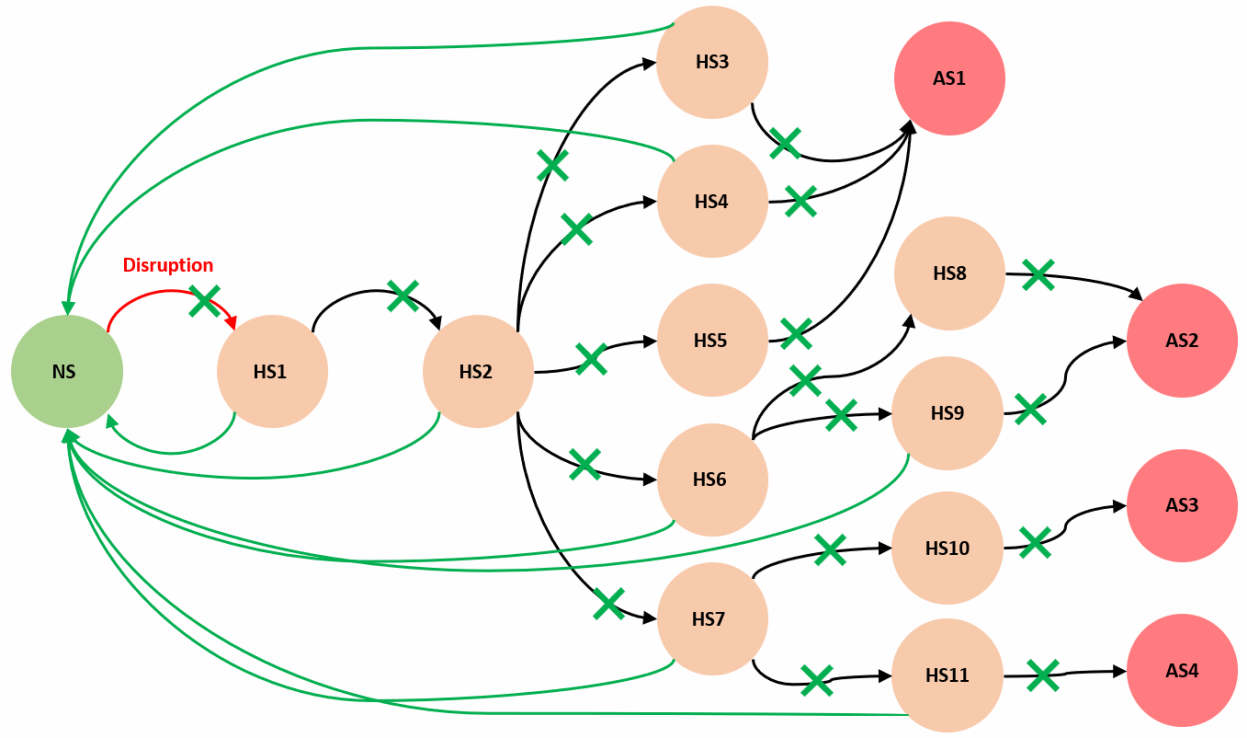

Figure 5.5: State and Trigger Model for *Micrometeorite Impacts Habitat*

In this example, a micrometeorite impacts the habitat outer shell, damaging the structural seals (**HS1**), a subsystem hazardous state. A habitat with damaged structural seals may also indicate a breach. Then, because the structural seals are a subsystem of the structural system, we identify the possible hazardous state that the structural system is not functioning properly (**HS2**), a system hazardous state. From this system hazardous state, we identify three possible habitat level hazardous states (**HS3**, **HS4**, and **HS5**), which describe either the habitat internal pressure is too low, is slowly decreasing, or is rapidly decreasing. If these habitat level hazardous states are left uncontrolled an accident state (**AS1**) of an unlivable pressure environment will ensue. Considering system dependencies, the structural thermal system is dependent on the structural system. Thus, we identify **HS6**, or the structural thermal system is not functioning properly, resulting from **HS2**. From **HS6** we identify two habitat level hazardous states **HS8** and **HS9** which are that the internal habitat pressure is either too low or too high. If these habitat level

hazardous states are left uncontrolled an accident state (**AS2**) of an unlivable temperature environment will ensue. The third system dependency on the structural thermal system is the radiation protection system, and a damaged structural system may cause a damaged radiation protection system (**HS7**). From this system level hazardous state, we identify two paths. The first is that the internal radiation level is above a livable condition (**HS10**) which, if left uncontrolled, may result in radiation poisoning (**AS3**). The second is that electric charge may build up on the exterior of the habitat due to the influx of radiation-induced charged particles (**HS11**) which, if left uncontrolled, may result in a loss of power due to electric discharge (**AS4**). In this example, we show the many logical paths that a disruption can take through hazardous states and eventually cause accident states. We design safety controls along these paths to avoid the propagation to hazardous states and eventually accident states.

We show in Figure 5.5 that the safety control designed to address the disruption prevents the propagation from the nominal state to the hazardous state. This is true for all safety controls designed to address disruptions. Also, the safety controls that prevent transition to either hazardous states or accident states (represented by the green crosses) are associated with the hazardous states that they are preventing the transition from. For example, the green crosses that prevent the transition from **HS2** to **HS3**, **HS4**, **HS5**, **HS6**, and **HS7** are all the same safety control *Ability to isolate the breached part of the habitat*, designed for **HS2**. When we design safety controls for disruptions, we are preventing the propagation of the habitat system from the nominal state to the hazardous state, we are not preventing the disruption from happening. Additionally, when we design safety controls for hazardous states, we assume that the hazardous state has already been reached. We use safety controls to return the habitat back to the nominal state or to prevent further hazardous or accident states. Finally, the triggers that are not disruptions (represented by the black arrows) are *No action* triggers.

Next, we convert this state and trigger model to the format we use in the failure network. There are several key differences in the way the data is presented in the failure network compared to the state and trigger model. These differences are presented in Table 5.5. Also, we implement a different labeling convention for the failure network.

Table 5.5: Differences in Data Representation between State and Trigger Model and Failure Network

Modeling Objective	**State and Trigger Model**	**Failure Network Format**
Disruptions	Disruptions are represented as triggers, or arrows, that transition the nominal state to a hazardous state.	Disruptions are represented as nodes. The nominal state node links to the disruption node, which then links to a hazardous state node.
Hazardous states	Hazardous states are represented as orange circles.	Hazardous states are represented as red nodes. This is purely a visualization choice, and the colors can be changed as preferred.
No action triggers	*No action* triggers are black arrows that transition a hazardous state to another hazardous state, or a hazardous state to an accident state.	*No action* triggers are represented as links from hazardous state to hazardous state. We assume any link from a hazardous state to a hazardous state is a *No action* trigger.
Accident states	Habitat level hazardous states transition to accident states through triggers.	We do not currently include accident states in the failure network. By definition, if left uncontrolled, the habitat level hazardous states will transition to accident states. The network is capable of storing accident states, but the scope of this thesis does not cover including them in the network.
Safety controls that prevent transition between states	Safety controls that prevent transition from a nominal state to a hazardous state or from a hazardous state to other hazardous or an accident state are represented as green crosses.	Safety controls that prevent transition from a specified state are represented as safety control nodes that link directly to the state that they are preventing transition **from.**
Safety controls that return the habitat to a nominal state	Safety controls that return the habitat to a nominal state are represented as triggers, or arrows that transition to the nominal state.	Safety controls that return the habitat to a nominal state from a specified hazardous state are represented as nodes. The specified hazardous state links to the safety control node, and then the safety control node links to the nominal state node.

The converted state and trigger model for this example is shown in Figure 5.6. We include additional safety controls and note that the labeling convention is different than the state and trigger model.

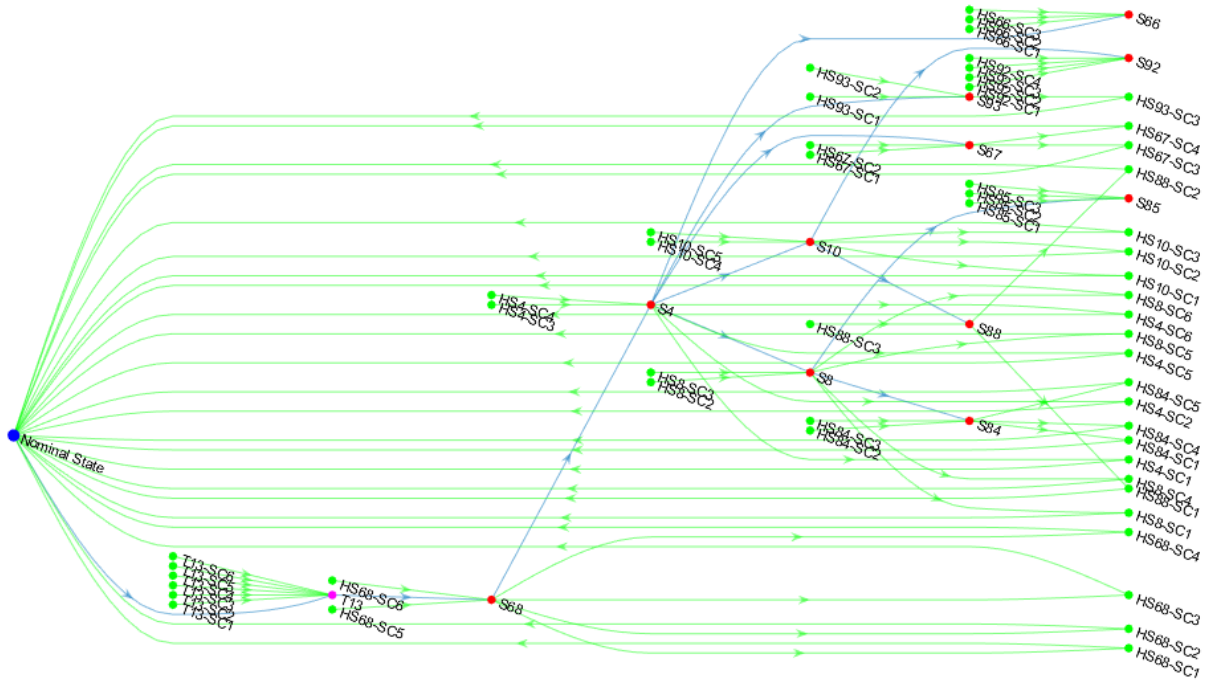

Figure 5.6: Network Format representation of the State and Trigger Model in Figure 5.5

In our failure network format, we show that we closely follow the structure of the state and trigger model. We aim to keep the structure of the failure network similar to the structure of the state and trigger model to allow for ease of comparison, and so that state and trigger models can be easily converted to into the network format. The nominal state (blue node) transitions through the disruption (magenta node) to a hazardous state (red node), which transitions other hazardous states, following the blue arrowed links. The safety controls are added using the green nodes and the green links, and either directly connect to the hazardous state or disruption they are designed to control or link from a hazardous state then to the nominal state. We stay consistent by noting that a disruption, denoted as a trigger in the state and trigger model, is represented as a node that links the nominal state to a hazardous state. Inversely, a safety control that returns the habitat to the nominal state, denoted as a trigger in the state and trigger model, is represented as a node that links a hazardous state to the nominal state. A key limitation of our current conversion of the state and trigger model to a network format is that all triggers that connect hazardous states to other hazardous states are *No action* triggers. This is not the case in reality, and we are working to improve the network format to include action triggers that we will code as appropriate. Once we have established consistency between the state and trigger model and the network format, we

repeat this process for all the state and trigger models that we create for each of the identified disruptions. The resulting failure network with added safety controls is basically the combination of all the state and trigger models converted to the network format following the guidelines in Table 5.5. We show the resulting failure network with added safety controls in Figure 5.7.

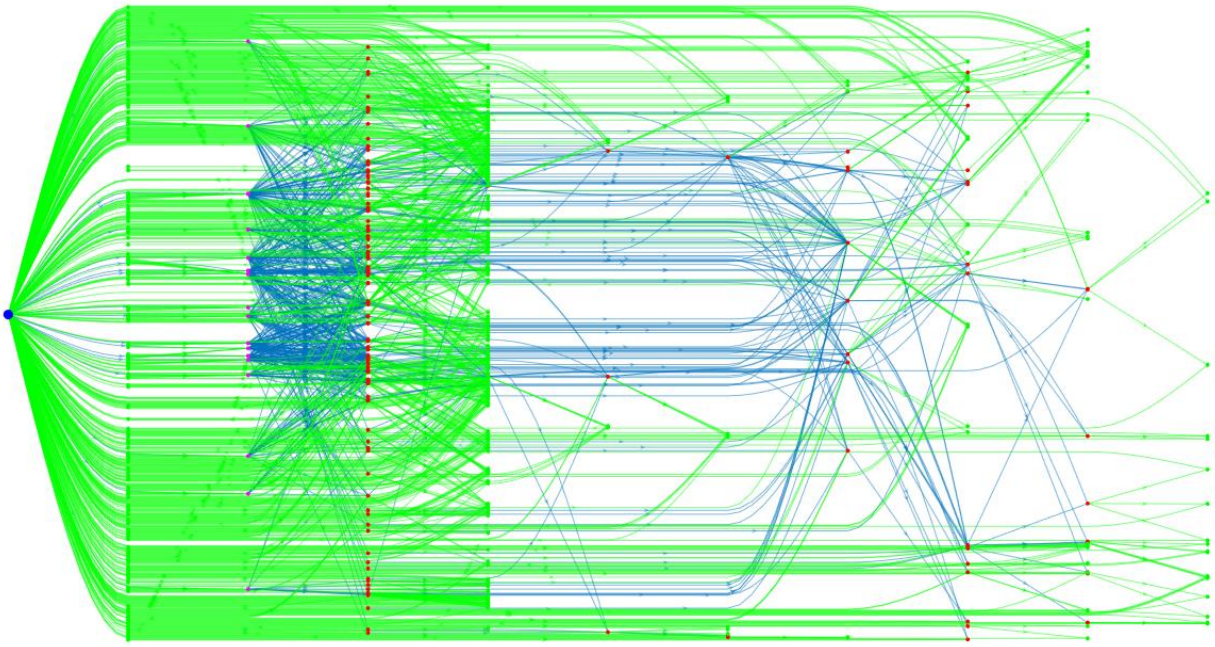

Figure 5.7: Failure Network with added Safety Controls

We combine all the state and trigger models into this format to fully realize the relationships between the different disruptions and hazardous states, hazardous states to hazardous states, and safety controls to both hazardous states and disruptions. Figure 5.7 shows all 776 identified safety controls linked to the hazardous state or disruption that they are designed to address. This method of network visualization can not only allow for a high-level analysis of the connection of different disruptions, hazardous states, and safety controls, it can also be a tool for isolating one disruption or hazardous state and considering the safety controls needed. For example, we showed this isolation technique in Figure 5.2 and Figure 5.6. As we develop the database and explore the relationships between the nodes, we can use different network theory techniques to measure the impact of each node within the network and use this information to direct efforts for creating safety controls to ensure system resilience.

6. ASSESSING THE EFFECTIVENESS OF OUR SAFETY CONTROLS TO MITIGATE THE DISRUPTIONS AND HAZARDOUS STATES

In this chapter, we discuss how we identify and use safety control flaws to identify how controls may be or become ineffective, and how we use safety control flaws as part of the development of the control effectiveness metric. This chapter covers step 4 in our control-theoretic approach to resilient design (see Figure 2.3), and steps 4a and 4b of our process for identifying safety controls (see Figure 3.1). Lastly, we apply our control effectiveness metric to an example disruption and an example hazardous state from our database.

6.1 Application: Safety Control Flaws and Generic Safety Control Flaws for a Martian Habitat

Steps 4a and 4b: Develop safety control flaws and generic safety control flaws

The final steps in the design for resilience process are to identify safety control flaws and generic safety control flaws. Control flaws are how safety controls may be or become ineffective, and we identify control flaws so that we can mitigate them by improving the safety controls or layering on additional safety controls. We cover safety control flaws in this chapter because we use them in the development of the control effectiveness metric, described in the next sections.

To identify safety control flaws, we organized each safety control by its associated generic safety control and determined how each safety control could fail or not fulfill its purpose. We asked, "how could this safety control break?". For example, the first generic safety control in our database is *ROBOT REPAIRS COMPONENT*. An example of a safety control could be a structural repair of the habitat or fixing a valve or sensor. To identify control flaws associated with this safety control, we found ways in which a robot may not be able to complete the repair that it was tasked to do. For example, the robot could simply not complete the repair, the robot could complete the repair, but imperfectly, the robot could damage the component or system while completing the repair, the robot could damage something else while making the repair, or the robot could do the repair too slowly or start too late. So far in our analysis we have designed safety controls as if they were perfectly designed, implemented, and operated. Identifying safety control flaws helps us understand where we may need to reinforce some safety controls and where, if needed, safety

controls are simply not enough to address a certain disruption or hazardous state. Another example of a generic safety control in our database is *COMPONENT WITHSTANDS SOURCE*. Identified safety control flaws could be that the component is simply not capable of withstanding the source, the component is not strong enough to withstand the source, or the component does not withstand the source sufficiently. Another generic safety control is *COMPONENT CORRECTS FOR SOURCE*. Identified safety control flaws could be that the correction controls the source but not for as long as needed, the correction is not activated completely or on time, the safety control corrects too much for the source, or the correction does not completely control the source. After completing this exercise for all 17 of our current generic safety controls, using examples of safety controls from each generic safety control to determine safety control flaws, we created a list of 64 safety control flaws. Like the safety controls and generic safety controls, we then identified the principle of the safety control flaws to create generic safety control flaws.

Our control theoretic approach to resilience relies on the principle that the cause of an accident is viewed as the result of a lack of constraints (or safety controls) imposed on the system design and operation rather than a series of events (Leveson, 2004). When safety controls are enforced inadequately, hazardous states and further accident states occur. Leveson, in her work developing the STAMP (System Theoretic Accident Model and Processes) accident model, developed a classification of accident factors that involve inadequate enforcement of safety constraints. Leveson's STAMP model is based on the hypothesis that accidents result from inadequate enforcement or control of safety-related constraints on the design, development, and operation of a system. In a similar way in which we use our safety controls to maintain safe behavior by addressing disruptions and hazardous states, in STAMP the goal of the control structure is to maintain a system, conceptualized as a continually adapting dynamic process, in a state of dynamic equilibrium enforced by feedback loops of information and control. She theorizes that "unsafe behavior results from either a missing or inadequate constraint or inadequate enforcement of the constraint leading to its violation". Thus, classification of control flaws starts by examining each control's potential contribution to inadequate control. Leveson identified three types of inadequate control:

(1) the controller may issue inadequate or inappropriate control actions,

(2) control actions may be inadequately executed, or

(3) there may be missing or inadequate feedback.

Each of these classifications can be further broken down for use in the STAMP model, and we focus on these three methods of inadequate control to develop our generic safety control flaws. Our generic safety control flaws involve the implementation of inadequate safety controls, the inadequate execution of safety controls, and possible missing or inadequate feedback from safety controls. When a safety control is not issued correctly, is inadequately executed, or does not provide feedback, we define that safety control as an unsafe control action. Leveson identifies four ways that unsafe control actions can occur:

(1) A safe control action is not provided,

(2) an unsafe control action is provided,

(3) a safe control action is provided too late or too early, and

(4) a safe control action is stopped too soon or applied too long.

Using these generic classifications as a starting point for our generic safety control flaws, we mapped our 64 safety control flaws to each of these generic safety control flaws. When an identified safety control flaw did not follow the principle of an original generic safety control flaw, we identified that safety control flaw principle and created a new generic safety control flaw to expand our list of generic safety control flaws. This process is similar to the identification of generic safety controls and safety controls, and after repeating it for all the identified safety control flaws, we determined six total generic safety control flaws. Table 6.1 shows each of these generic safety control flaws along with a description.

Table 6.1: Generic Safety Control Flaws

Generic Safety Control Flaw	**Description**
SAFE CONTROL ACTION IS NOT PROVIDED	Describes when the safety control is not implemented for any reason. This could be for example because it is not possible to be implemented, it is chosen not to be implemented, or that the safety control was attempted but not completed successfully.
SAFE CONTROL ACTION IS PROVIDED TOO LATE OR TOO EARLY	Describes when the safety control was stopped too soon or was activated too quickly. The safety control may not be completed in time for example because of a long or complicated procedure, lack of autonomous action, or lack of available resources to complete the safety control. A safety control may also be implemented too quickly, in that for example a component may be replaced before it needs to be. This flaw highlights the need for appropriate timing for implementing the safety control.
SAFE CONTROL ACTION IS PROVIDED TOO MUCH OR TOO LITTLE	Describes when the safety control was not adequate to protect against the source, or when the safety control provides too much protection against the source that it becomes detrimental to other parts of the habitat. For example, shielding may be inadequate to protect against a micrometeorite, or crew protection may not be enough to protect against radiation. Also, a safety control may redirect power or use redundant systems to correct an underperforming system taking away resources and power from other systems that may need it.
SAFE CONTROL ACTION CAUSES UNSAFE CONTROL ACTION	Describes when the safety control execution makes the current hazardous state worse, as in that safe control action ends up causing an overall unsafe control action. For example, a repair could be completed incorrectly. A good intentioned repair, or a safe control action, is completed incorrectly and the component or system performs worse than before, creating an overall unsafe control action
SAFE CONTROL ACTION CAUSES HAZARDOUS STATE	Describes when the safety control execution leads to other hazardous states. For example, if a robot is completing a repair on the exterior of the habitat and becomes damaged by falling from the repair site, that robot is now in a hazardous state. Another example is that if a component or system is electrically isolated because it is malfunctioning, disrupting the current to that system may cause the systems in that circuit to malfunction as well.
SAFE CONTROL ACTION IS APPLIED TOO LONG OR STOPPED TOO SOON	Describes when the safety control is executed for an unnecessarily long period, or if the safety control is stopped prematurely in the event of a source of a disruption. For example, if the crew is relocated due to a dust storm and there is no indication of when the dust storm ends, the safety control will still be implemented and it will be implemented for too long. Adversely, if the crew exits the relocation area during the dust storm, that will not constitute a safe control action because it will have been stopped too soon.

These generic safety control flaws are used to develop part of the control effectiveness metric, covered in the next sections.

6.2 Developing the Control Effectiveness Metric

The control effectiveness metric is a collection of data that we use to discern how effective a safety control is at mitigating its target hazardous state or disruption. The metric contains information on the control's susceptibility to flaws, expected probability of success, availability, and competence

against a source of a disruption, and the expected time it would take to implement the control. Control effectiveness is intended to indicate how well a certain safety control addresses the hazardous state it was designed for. This metric can then be used as part of the overall system design process to select safety controls that contribute to a design with the desired performance and resilience. As we develop more safety controls for our space habitat, described in Chapter 5, we develop a large set of safety controls that may or may not be effective in controlling their associated disruption or hazardous state. This section describes how we discern which safety controls may be worth pursuing further from the large set that we have already established. We use the control effectiveness of a safety control to down select our safety control option space and develop our *selected safety controls*, which are the safety controls we consider further for implementation into the habitat. These selected safety controls enter a trade space along with other habitat considerations, like performance and cost, to provide guidance to designers of resilient architecture. The following procedure outlines how we use the control effectiveness metric in the context of our failure network and safety control option space.

First, we choose a disruption, subsystem level hazardous state, system level hazardous state, or habitat level hazardous state to control based on our hazardous state and disruption assessment techniques covered in Chapter 4. Once we choose a disruption or hazardous state, we use our safety control option space to identify the list of safety controls for the chosen disruption or hazardous state.

Up to this point, the safety controls have been focused only on *what* needs to be done to prevent propagation to further hazardous states or return the habitat to the nominal state. Many of the safety controls begin with the phrase *Ability to…*, but to develop a better understanding of the effectiveness of a safety control we need to understand *how* the safety control achieves its control goal. For example, we choose a safety control *Ability for solar arrays to vacuum or move dust*. There are several ways to achieve this safety control function. We could implement a built-in vacuum in the solar arrays that automatically removes dust, we could task a robot agent to brush off the dust or vacuum the dust, or similarly we could task a human agent to brush off the dust or vacuum the dust. For the identified safety control, we create several **implementation strategies** that specify how the safety control goal will be achieved. Once we have identified the implementation strategies for a selected safety control, we then want to know how effective each

implementation strategy is. To address this, we ask the following questions: (1) If the implementation strategy is perfectly implemented, how well does it work? (2) If the implementation strategy is not perfectly implemented, what impact will it have on the component or system? (3) If the implementation strategy is hazardously implemented, what impact will it have on the habitat?

We address the first question by creating the metric *implementation strategy effectiveness (ISE)*, a precursor to control effectiveness. Within this metric, we define four values that seek to answer how well the implementation strategy works assuming it is perfectly implemented. The first value, a probability from 0 to 1, is the probability that the implementation strategy will be implemented perfectly. For example, a complex or multi-step implementation strategy may have a lower probability of being implemented perfectly than a simple or straight forward implementation strategy. The second value, also a probability from 0 to 1, is the probability that the implementation strategy is available at the time of control. Factors that affect the probability of availability include the possibility that the implementation strategy is only available at certain times, or that the implementation strategy is disposable and can only be used once or a few times. An implementation strategy with a higher probability of availability is more favorable and that will factor into its effectiveness. The third value, again a probability from 0 to 1, is the probability that the implementation value will successfully control the source, or its probability of competence. This is like the probability that the implementation strategy will be implemented perfectly but differs in the sense that even if the implementation strategy is implemented perfectly it may still not successfully control the source. For example, an implementation that, if implemented perfectly, successfully controls the source 70% of the time, will have a probability of competence of 0.7. The final value contained in the implementation strategy effectiveness metric is the active control time. This value is measured in units of time and could range from seconds to days. A shorter active control time is more favorable in terms of implementation strategy effectiveness. The metric is shown as a collection of these probabilities and active control time, in Equation 6.1.

$$ISE = [P_{perfect}, P_{available}, P_{competent}, t_{active}]$$

Equation 6.1: Implementation Strategy Effectiveness

For questions (2) and (3), we use our generic safety control flaws. For each implementation strategy of a given safety control, we determine the negative impact the implementation strategy has if it is imperfectly implemented or hazardously implemented by assessing the application of each of our previously identified generic safety control flaws. An implementation strategy is imperfectly implemented if, for example, it is simply not provided, or it is provided too late or too early, too much or too little, or applied too long or stopped too soon. An implementation strategy is hazardously implemented if, for example, the strategy causes a hazardous state, or it causes an unsafe control action. We have gathered these common flaws of safety controls in Chapter 5, and we use them in our implementation strategy effectiveness and further in control effectiveness. To do this, we draw inspiration from the risk priority number (RPN) used in failure mode effects analysis (FMEA). RPN is usually presented as the product of three measures: severity, occurrence, and detection, each assigned a score from 1, least, to 10, best. Although this risk assessment strategy is used to assess failure and failure modes in a system and their causes and effects, we specifically repurpose the risk priority number for our control effectiveness metric. For each implementation strategy of each safety control, we assess the negative impact on the component, system, or habitat using the generic safety control flaws. For each generic safety control for a given implementation strategy, we assign a severity score and an occurrence score. For our purposes, we use a rating system consistent with the system NASA has adopted and has used recently on the Orion Crew Exploration Vehicle. Both the likelihood and severity ratings range from 1 (very low), to 5 (very high). For example, an event with a likelihood rating of 5 has a qualitative description, in that it is "likely to occur", and has a quantitative description, in that the probability is greater than 50% that this event will happen (Perera, 2012). Thus, in the case that an implementation strategy will exhibit a particular generic control flaw over 50% of the time, we rate that likelihood as a 5. Similarly, an event that may warrant a severity rating of 5 would cause a condition that may cause death or loss of crew, cause destruction of critical facilities or vehicle, or cause termination of the project. The product of those scores for a given generic safety control flaw for a given implementation strategy is the criticality that is associated with that strategy in respect to that flaw. When we sum the criticality scores for a given implementation strategy, and we result in an indication of how flawed that implementation strategy may be and how critical it will be to the system or habitat if the flaws are not addressed. The criticality equation for a given implementation strategy is shown in Equation 6.2.

$$Criticality_{IS} = \sum_{i=1}^{n} Likelihood_{IS,i} * Severity_{IS,i}$$

Equation 6.2: Criticality of an Implementation Strategy with respect to Generic Control Flaw i

We do this process for all implementation strategies and for all generic safety control flaws, and we include these criticality scores alongside implementation strategy effectiveness. Thus, for a given hazardous state, we create Table 6.2.

Table 6.2: Structure of Implementation Strategy Metrics for a given Hazardous State with two Safety Controls

Hazardous State			
Safety Control 1		Safety Control 2	
$criticality_{1,1}$	$ISE_{1,1}$	$criticality_{2,1}$	$ISE_{2,1}$
$criticality_{1,2}$	$ISE_{1,2}$	$criticality_{2,2}$	$ISE_{2,2}$
$criticality_{1,3}$	$ISE_{1,3}$	$criticality_{2,3}$	$ISE_{2,3}$

With the information in Table 6.2, we create the basis for control effectiveness. For safety control 1 in Table 6.2, control effectiveness is a data structure containing the criticality and the implementation strategy effectiveness for each implementation strategy considered for safety control 1. Similarly, for safety control 2, control effectiveness is a data structure containing the criticality and the implementation strategy effectiveness for each implementation strategy considered for safety control 2. We use this method to give the designer freedom to decide which safety control to use and which implementation strategy to use in an informed, systematic, and tiered approach. First, the designer considers each safety control at face value, considering all the possible implementation strategies and criticalities. If the implementation strategies for a given safety control tend to have a high effectiveness and a low criticality, then we can consider that safety control further. If the opposite is true, we may decide to not further pursue that safety control because all the implementation strategies considered for that safety control are not sufficient to control the source. If the implementation strategies for a given safety control vary in terms of effectiveness and criticality, we can investigate each implementation strategy and choose a strategy based on which have a high effectiveness, a low criticality, or both. This method of organizing control effectiveness aims to answer the following questions: (1) Which safety control is the least likely to be flawed? (2) How should we implement that safety control? And (3) Based on all the

safety control and considered implementation strategies for this hazardous state or disruption, what control has the potential to be the most effective?

To answer question (1), the designer looks at each criticality score for the implementation strategies for a given safety control. This will give the designer an idea of which safety controls are the least likely to have flaws, and how severe the implications will be if the safety control is in fact flawed. To answer question (2), the designer looks at the implementation strategy effectiveness and considers the probability of success and the associated values of active control time, probability of availability, and probability of competent control. Considering these values along with the criticality value aids in the trade-off process to decide in what way to implement the safety control. To answer question (3), the designer can consider all the data contained in each control effectiveness data structure for each safety control. This information can clarify the possible effectiveness of each safety control based on how it is implemented, without necessarily knowing the implementation strategy. Simply being presented with the data on each safety control, a designer can know how effective each safety control may be as well as how critical it may be if the safety control is flawed. This knowledge aids in the trade-off process to decide which safety control should be pursued further. To further aid in this process, we develop a color-coded grading scheme for each set of values in the control effectiveness data structure. Table 6.3 shows how we create a color-coding scheme for each value included in the control effectiveness metric. For criticality, the maximum criticality is 600 and the minimum is 10, based on the number of generic safety control flaws we have identified so far. To normalize this measure, we grade the criticality based on the possible minimum and maximum criticality score, accounting for the addition of further generic safety control flaws.

Table 6.3: Control Effectiveness Color Coding Scheme

Value	Units	Range
Criticality	N/A	$C < 0.33 * C_{max}$
		$0.33 * C_{max} \leq C \leq 0.67 * C_{max}$
		$C > 0.67 * C_{max}$
Probability of being implemented perfectly	Probability (0 to 1)	$P_{perfect} > 0.75$
		$0.25 \leq P_{perfect} \leq 0.75$
		$P_{perfect} < 0.25$
Probability of availability	Probability (0 to 1)	$P_{available} > 0.75$
		$0.25 \leq P_{available} \leq 0.75$
		$P_{available} < 0.25$
Probability of competence	Probability (0 to 1)	$P_{competent} > 0.75$
		$0.25 \leq P_{competent} \leq 0.75$
		$P_{competent} < 0.25$
Active Control Time	Time [s], [m], [hrs], [days]	$t_{active} \sim [s]\ OR\ [m]$
		$t_{active} \sim [hrs]$
		$t_{active} \sim [days]$

Additionally, we note that the user can select the color mapping according to their risk preference. For example, if a user was more risk averse, these probabilities and criticalities can be changed to reflect that. A "green" probability of perfect implementation, probability of availability, and probability of competence would be set to a higher value or a more stringent window of acceptance. The cut-off for the "green" criticality measure would also be lowered to reflect this risk averse nature. Table 6.4 shows how one may change the control effectiveness color coding scheme to reflect a more risk averse nature.

Table 6.4: Example "Risk Averse" Control Effectiveness Color Coding Scheme

Value	Units	Range
Criticality	N/A	$C < 0.20 * C_{max}$
		$0.20 * C_{max} \leq C \leq 0.40 * C_{max}$
		$C > 0.40 * C_{max}$
Probability of being implemented perfectly	Probability (0 to 1)	$P_{perfect} > 0.85$
		$0.35 \leq P_{perfect} \leq 0.85$
		$P_{perfect} < 0.35$
Probability of availability	Probability (0 to 1)	$P_{available} > 0.85$
		$0.35 \leq P_{available} \leq 0.85$
		$P_{available} < 0.35$
Probability of competence	Probability (0 to 1)	$P_{competent} > 0.85$
		$0.35 \leq P_{competent} \leq 0.85$
		$P_{competent} < 0.35$
Active Control Time	Time [s], [m], [hrs], [days]	$t_{active} \sim [s]$
		$t_{active} \sim [m]\ OR\ [hrs]$
		$t_{active} \sim [days]$

Using this color-coding scheme, we can further add value to Table 6.2 to give a visual representation of how effective each implementation strategy of each safety control may be. Using arbitrary values, like in Table 6.2, we create Table 6.5 as an example.

Table 6.5: Structure of Control Effectiveness with Color Coding, for an example Hazardous State with two Safety Controls

Hazardous State										
Safety Control 1						Safety Control 2				
ISE Values	**Criticality**	**ISE**				**Criticality**	**ISE**			
Imp. 1	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}
Imp. 2	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}
Imp. 3	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}	Criticality	$P_{perfect}$	$P_{available}$	$P_{competent}$	t_{active}

We use Table 6.4 to gain a holistic view of which safety controls have the potential to be the most or the least effective. For example, in Table 6.4, the first arbitrary implementation strategy for the first safety control has a green criticality measure and green and orange values for the implementation strategy effectiveness. As designers, we are more inclined to further investigate this implementation strategy for this safety control than the last implementation strategy for the second safety control, which shows mostly red and orange values for criticality and the implementation strategy effectiveness values. For the implementation strategies for the safety controls that show a range of green, orange, and red values, we can investigate these options further as part of a trade study that considers the benefits of each one in terms of risk and effectiveness. We have created a method in this section that designers can use to down select safety controls and the implementation of those safety controls based on each safety control's effectiveness in controlling a particular hazardous state, and also considering the risk involved in implementing certain safety controls based on our established safety control flaws. In the next sections, we apply this method to our example space habitat, using our database of disruptions, hazardous states, and safety controls.

6.3 Application: The Control Effectiveness Metric

In this section, we use our control effectiveness metric to measure the control of an example disruption and an example hazardous state. We used our disruption and hazardous state assessment techniques described in Chapter 4 to choose a disruption and hazardous state to control. We

consider ionizing radiation to have a large local importance in terms of severity and probability, and we showed that in our network format the disruption led to many subsystem level hazardous states. Thus, we will apply our control effectiveness metric to the ionizing radiation disruption. Additionally, a hazardous state that we showed to have a high severity and to have many documented relationships in the network format is *Oxygen concentration in the habitat is too low*. We consider that hazardous state as a test case for our control effectiveness metric in this section. We follow the procedure described in the previous section to assess implementation strategies for several safety controls that we have identified.

6.3.1 The Control Effectiveness Metric and an Example Disruption

To consider a disruption to control, we refer to our network measures and use established hazard analysis to identify a disruption that may be critical to operation of the habitat. An example disruption that leads to many subsystem level hazardous states is *Ionizing Radiation*. Additionally, ionizing radiation is a disruption that is continuously present in the habitat environment and has severe implications for the crew's well-being. Thus, radiation is a disruption that is both severe and probable and has a high network risk criticality within our failure network.

First, we identify the safety controls that we have created for the ionizing radiation disruption. Table 6.5 shows these safety controls.

Table 6.6: Example Safety Controls for Ionizing Radiation Disruption

Disruption ID	**Disruption Description**	**Disruption Risk Level**
T5	Ionizing Radiation	Critical

Safety Control ID	**Safety Control Description**	**General Safety Control**
T15-SC1	Provide parts of the habitat that have more radiation protection for when radiation spikes	REMOVE COMPONENT FROM SOURCE
T15-SC2	Provide individualized crew radiation protection during spikes in radiation activity	EXTRA PROTECTION FOR HUMANS FROM SOURCE
T15-SC3	Bury the habitat under layers of regolith to avoid radiation	REDUCE COMPONENT LOAD

For each one of these safety controls, we create implementation strategies. In this example, we will create three implementation strategies for each of the three safety controls to form a control decision matrix, however we note that the number of implementation strategies is not fixed. The first safety control, *Provide parts of the habitat that have more radiation protection for when radiation spikes*, makes available a "safe room" concept where the crew can shelter in a specific

part of the habitat when radiation levels rise to a potentially dangerous level. Although radiation is continuously present, solar events may cause radiation to fluctuate deeming necessary extra radiation protection during these events. The first implementation for a "safe room" concept would be to strategically place the water reserves in the habitat to shield one compartment of the habitat from radiation. The water reserves are needed for long term crew habitation, and in this implementation, we take advantage of the excellent radiation particle blocking ability of water to place the water reserves in the walls of the habitat surrounding a specific room. The water can then be periodically replaced with filtered wastewater using the habitat's water management system, part of the ECLSS. Another implementation for this safety control would be to use hydrogenated boron nitride nanotubes (BNNTs) that are excellent shields for radiation. These tiny nanotubes made of carbon, boron, and nitrogen, with hydrogen interspersed throughout the empty spaces left in between the tubes, are excellent absorbers of radiation and can be woven into structure and fabric (Frazier, 2017). The third implementation that we consider for this safety control is to implement a small, localized electric field that creates a protective bubble around a compartment of the habitat to shield from radiation. Like the Earth's magnetic field protects inhabitants from energetic particles, an artificial electric or magnetic field could create a protective bubble around a spacecraft or habitat (Frazier, 2017). These implementation strategies are examples of how to implement the safety control of providing parts of the habitat that have more radiation protection for when radiation spikes. We create the safety control to specify what we should do to mitigate the disruption, and the implementation strategies help define how we should mitigate the disruption. These implementation strategies could be added to with the aid of radiation experts, and this process is necessarily collaborative. For our purpose, we aim to demonstrate the process to implement the control effectiveness metric. We continue to create implementation strategies for the other two specified safety controls for this disruption, and Table 6.6 shows the safety controls with their respective implementation strategies for the mitigation of ionizing radiation.

Table 6.7: Safety Controls and Implementation Strategies for Ionizing Radiation Disruption

Disruption ID	**Disruption Description**	**Disruption Risk Level**	
T5	Ionizing Radiation	Critical	
Safety Control:	Provide parts of the habitat that have more radiation protection for when radiation spikes	Provide individualized crew radiation protection during spikes in radiation activity	Bury the habitat under layers of regolith to avoid radiation
Implementation 1:	Strategically place water reserves to shield one compartment of the habitat from radiation, and periodically replace the water with filtered wastewater	Implement radiation suits available for crew to wear when inside the habitat	Implement in-situ resource utilization to place layers of regolith on top of the already built habitat
Implementation 2:	Implement hydrogenated boron nitride nanotubes in one compartment of the habitat to shield from radiation	Implement space suits available for crew to wear to avoid excessive radiation when inside the habitat	Build the habitat partially underground so that some of the habitat is protected from radiation spikes
Implementation 3:	Implement a small, localized electric field that creates a protective bubble around a compartment of the habitat to shield from radiation	Administer radiation medication that diminished the effect of radiation on the crew	Build the habitat completely underground to avoid radiation effects

To demonstrate our application of the control effectiveness metric for this example, we consider the values shown in Table 6.3 and we apply them to an example implementation strategy of a safety control. We consider the first implementation strategy of the first safety control: *Strategically place water reserves to shield one compartment of the habitat from radiation, and periodically replace the water with filtered wastewater*. The first value of the implementation strategy effectiveness, part of the larger control effectiveness exercise, is criticality. We obtain criticality by considering the implementation's susceptibility to the generic safety control flaws that we identified in Section 6.1. The more the generic control flaws apply to the implementation, the higher the criticality, and the greater the risk associated with implementing that strategy becomes. We consider both the likelihood and severity associated with the implementation strategy eliciting a control flaw and multiply those numbers together. For example, our first generic control flaw, *SAFE CONTROL ACTION IS NOT PROVIDED*, describes the possibility that the safety control is not implemented for any reason. This could be for example because it is not possible to be implemented, it is chosen not to be implemented, or that the safety control was attempted but not completed successfully. For the first implementation of the first safety control, the likelihood of not having the appropriate water stores in a strategic position to block the radiation particles, as

well as not being able to contain the crew in a specific compartment in the habitat before being exposed to excessive radiation is high. Because of these limitations, we assign a score of 3 out of 5 to the likelihood that this implementation strategy elicits this control flaw. Additionally, the severity of not providing this implementation strategy is also high, as that would expose the crew to excessive radiation. Thus, we assign a score of 4 out of 5 to the severity of this implementation not being provided. Multiplying these scores together, we obtain a criticality of 12 for the first implementation of the first safety control eliciting the first generic safety control flaw. We complete this exercise for the remaining generic control flaws for this implementation strategy, and we sum the criticality score for each generic control flaw and obtain a total criticality score of 82. In Table 6.3 we defined a color-coded scheme for a criticality that is considered low, medium, or high. In this case, we consider the criticality to be medium, and assign an orange color.

Additionally, we consider the probabilities contained in the implementation strategy effectiveness for the first implementation strategy. The first probability we consider is the probability that the implementation is implemented perfectly. For this implementation to be implemented perfectly, the compartment contained with water shielding must be completely contained, must have a sufficient thickness of water to block the incoming radiation, the water must not be contaminated with particles that may interfere with the radiation protection capabilities, and the crew must be able to quickly contain themselves in the compartment. Although these requirements seem stringent, it is not unrealistic to think that the crew would be able to satisfy these requirements. Water is needed for operation of the habitat and can be created through hydrolysis, and the compartment walls can be made thick and stable enough to accommodate extra water storage. For these reasons, we assign a 0.7 probability that this implementation is implemented perfectly. According to the color-coding scheme defined in Table 6.3, this is a medium probability. The second probability that we assign as part of the implementation strategy effectiveness is the probability that this implementation will be available for use. We maintain that water will be a necessity in the habitat, and the crew should never have a water shortage. However, in the order of necessity, the crew should be able to consume the water rather than use it for radiation protection. Thus, we propose that there is a high likelihood that water will be available for radiation protection, but we must maintain that water will not always be available for radiation protection. We assign a 0.8 probability that this implementation will be available for use. The third and final probability that we consider as part of the implementation strategy effectiveness is the probability that the

implementation strategy will be competent in its control of the disruption. That is, all other considerations aside, how competent is this strategy at protecting against the source of the disruption? Since, as stated, water is an excellent blocker of radiation particles, we assign a probability of competence of 0.9, a high competence level. Finally, we consider the time this control takes to mitigate the source. For this, we consider how long it may take the crew to relocate to a specific compartment, we consider that if water is not available in the stores how long it may take to pump water into the walls of the compartment, and we consider how long it may take the crew to jointly recognize that there is a spike in radiation activity around the habitat. Thus, we consider that this implementation strategy has a timescale of minutes, but no longer than ten minutes. This is an approximation, but we consider that this implementation strategy can be implemented relatively quickly, and assign a good, or green, timescale.

We repeat this process of assigning criticality scores, probabilities, and timescales for all the implementation strategies of the safety control option space for this disruption and create the risk decision matrix shown in Figure 6.1.

Disruption:	Ionizing Radiation														**Critical**
Safety Control:	Provide parts of the habitat that have more radiation protection for when radiation spikes					Provide individualized crew radiation protection during spikes in radiation activity					Bury the habitat under layers of regolith to avoid radiation				
Implementation 1:	Strategically place water reserves to shield one compartment of the habitat from radiation, and periodically replace the water with filtered wastewater					Implement radiation suits available for crew to wear when inside the habitat					Implement in-situ resource utilization to place layers of regolith on top of the already built habitat				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	82	0.7	0.8	0.9	~ [m]	108	0.9	0.9	0.6	~ [m]	74	0.4	0.8	0.9	~ [days]
Implementation 2:	Implement hydrogenated boron nitride nanotubes in one compartment of the habitat to shield from radiation					Implement space suits available for crew to wear to avoid excessive radiation when inside the habitat					Build the habitat partially underground so that some of the habitat is protected from radiation spikes				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	46	0.8	0.8	0.9	~ [s]	96	0.9	0.9	0.7	~ [m]	52	0.4	0.8	0.8	~ [days]
Implementation 3:	Implement a small, localized electric field that creates a protective bubble around a compartment of the habitat to shield from radiation					Administer radiation medication that diminishes the effect of radiation on the crew					Build the habitat completely underground to avoid radiation effects				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	132	0.5	0.2	0.8	~ [m]	128	0.9	0.4	0.3	~ [m]	44	0.3	0.8	0.9	~ [days]

Figure 6.1: Risk Decision Matrix for *Ionizing Radiation* Disruption

Using Figure 6.1, we can actively select controls and implementations of those controls that are less likely to be flawed, have high probabilities of success, availability, and competence, and can be completed quickly. Control effectiveness considers all these measures, and at first glance we can select an implementation of a safety control that exhibits high, or green, performance in each of these categories. Thus, we can further consider the second implementation of the first safety control, which is implementing hydrogenated boron nitride nanotubes in the shell of the habitat

structure. The data contained in Figure 6.1 makes up the control effectiveness metric, and aids in the trade-off process for the designer to decide which safety control or controls should be pursued further. For example, using regolith to either cover the habitat exterior or to bury the habitat partially or fully underground is an excellent strategy at blocking radiation. However, these strategies are time consuming and may lead to negative consequences during construction and can lead to structural deformation. Additionally, implementing radiation suits or space suits for the crew to use while inside the habitat are quick, effective, and available solutions, but they may not be sufficient as long-term solutions for radiation impact. We can modify our decision-making process by implementing more stringent guidelines on the values in our control effectiveness metric, reflecting the risk averse perspective shown in Table 6.4. By doing this, we create the risk decision matrix in Figure 6.2.

Disruption:	Ionizing Radiation														**Critical**
Safety Control:	Provide parts of the habitat that have more radiation protection for when radiation spikes					Provide individualized crew radiation protection during spikes in radiation activity					Bury the habitat under layers of regolith to avoid radiation				
Implementation 1:	Strategically place water reserves to shield one compartment of the habitat from radiation, and periodically replace the water with filtered wastewater					Implement radiation suits available for crew to wear when inside the habitat					Implement in-situ resource utilization to place layers of regolith on top of the already built habitat				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	82	0.7	0.8	0.9	~ [m]	108	0.9	0.9	0.6	~ [m]	74	0.4	0.8	0.9	~ [days]
Implementation 2:	Implement hydrogenated boron nitride nanotubes in one compartment of the habitat to shield from radiation					Implement space suits available for crew to wear to avoid excessive radiation when inside the habitat					Build the habitat partially underground so that some of the habitat is protected from radiation spikes				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	46	0.8	0.8	0.9	~ [s]	96	0.9	0.9	0.7	~ [m]	52	0.4	0.8	0.8	~ [days]
Implementation 3:	Implement a small, localized electric field that creates a protective bubble around a compartment of the habitat to shield from radiation					Administer radiation medication that diminishes the effect of radiation on the crew					Build the habitat completely underground to avoid radiation effects				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	132	0.5	0.2	0.8	~ [m]	128	0.9	0.4	0.3	~ [m]	44	0.3	0.8	0.9	~ [days]

Figure 6.2: “Risk Averse” Risk Decision Matrix for *Ionizing Radiation* Disruption

Using these measures, we aid in the design process to assist in the decision of which solutions most effectively mitigate the disruption in question. We also can consider implementing multiple controls to counteract drawbacks of some safety controls with strengths of others, for example the crew could wear radiation suits until the habitat is sufficiently buried under regolith. From a safety control perspective, we use control effectiveness to contribute design choices to support a resilient habitat architecture. These design choices should be fed into a larger design space, considering other system requirements such as cost, performance, mass, and crew psychological impacts.

6.3.2 The Control Effectiveness Metric and an Example Hazardous State

In this section, we similarly implement the control effectiveness metric that we completed for a disruption, but for a habitat level hazardous state. Using our network measures in addition to established hazard assessment techniques covered in previous sections, we identify a habitat level hazardous state: *Oxygen concentration in habitat is too low*. We cannot adequately estimate the probability of this hazardous state occurring, but we can estimate from our network measures that many disruptions and hazardous states can eventually lead to this habitat level hazardous state. Thus, we consider this hazardous state to be critical to control.

First, we identify the safety controls that we have designed to mitigate this hazardous state. These safety controls are shown in Table 6.8.

Table 6.8: Example Safety Controls for Low Oxygen Concentration Hazardous State

Hazardous State ID	**Hazardous State Description**	**Hazardous State Risk Level**
HS73	Oxygen concentration in habitat is too low	Critical

Safety Control ID	**Safety Control Description**	**General Safety Control**
HS73-SC1	Ability to provide oxygen masks and containers to the crew	EXTRA PROTECTION FOR HUMANS FROM SOURCE
HS73-SC2	Ability to isolate oxygen supply to specific areas of the habitat	ISOLATE COMPONENT/SYSTEM
HS73-SC3	Ability to increase output of oxygen generation system	COMPONENT CORRECTS FOR SOURCE

Repeating the process from the previous section, we create implementation strategies for each of these safety controls. We create three implementation strategies for each of these safety controls to satisfy the example, however the number of implementation strategies is not fixed. For the first safety control, *Ability to provide oxygen masks and containers to the crew*, we offer three implementation strategies to achieve this safety control. First, we can make available portable oxygen tanks and masks that the crew can use to maintain normal oxygen levels and still be able to get around the habitat. This is a rather obvious implementation of this safety control but is a necessary contingency to ensure that there is always oxygen available. The second implementation strategy for the first safety control is to implement reserve oxygen storage tanks, possibly embedded into the habitat structure, that drop oxygen masks from the ceiling or walls for the crew to use. This is a similar strategy used by airlines in case of emergency when the cabin of an aircraft loses pressure. An evident drawback of this strategy is that the crew will not be very mobile when using these oxygen stores, but this implementation strategy will most likely allow for more oxygen

storage than in portable tanks. The third implementation strategy for the first safety control is to have the crew wear their spacesuits, that have built in oxygen stores, until the oxygen levels have been restored. We complete this exercise for all three of the considered safety controls, and the safety controls and their implementation strategies are shown in Table 6.9.

Table 6.9: Safety Controls and Implementation Strategies for Low Oxygen Concentration Hazardous State

Hazardous State ID	**Hazardous State Description**	**Hazardous State Risk Level**
HS73	Oxygen concentration in habitat is too low	Critical

Safety Control:	Ability to provide oxygen masks and containers to the crew	Ability to isolate oxygen supply to specific areas of the habitat	Ability to increase output of oxygen generation system
Implementation 1:	Bring and store portable oxygen tanks that the crew can use to maintain normal oxygen levels and still be able to get around the habitat	Implement that at least one compartment of the habitat has the required minimum amount of oxygen concentration in the air, even if it means other rooms have insufficient oxygen	Implement a reserve oxygen generation system in addition to the nominal oxygen generation system
Implementation 2:	Implement reserve oxygen storage tanks and drop oxygen masks from the ceiling or walls for the crew to use	Modularize the habitat oxygen supply so that a failure in one compartment only affects the oxygen concentration in that compartment, and the crew can continue in other rooms	Implement oxygen candles in addition to the nominal oxygen generation system
Implementation 3:	Implement that the crew wear their spacesuits until the oxygen levels have been restored	Implement multiple, physically separated habitats so that a failure of the oxygen generator in one habitat does not mean the other habitat(s) is unlivable	Provide more power to the oxygen generation system so that the chemical reaction to produce oxygen is increased

To demonstrate our application of the control effectiveness metric for this example, we consider the values shown in Table 6.3 and we apply them to an example implementation strategy of a safety control. We consider the first implementation strategy of the first safety control: *Bring and store portable oxygen tanks that the crew can use to maintain normal oxygen levels and still be able to get around the habitat*. The first value of the implementation strategy effectiveness, part of the larger control effectiveness exercise, is criticality. We obtain criticality by considering the implementation's susceptibility to the generic safety control flaws that we identified in Section 6.1. The more the generic control flaws apply to the implementation, the higher the criticality, and the greater the risk associated with implementing that strategy becomes. We consider both the

likelihood and severity associated with the implementation strategy eliciting a control flaw and multiply those numbers together. For example, our first generic control flaw, *SAFE CONTROL ACTION IS NOT PROVIDED*, describes the possibility that the safety control is not implemented for any reason. This could be for example because it is not possible to be implemented, it is chosen not to be implemented, or that the safety control was attempted but not completed successfully. For the first implementation of the first safety control, the likelihood that the crew would not be able to implement the portable oxygen tanks and masks is relatively low. If available, it would not take much effort for the crew to find the oxygen tanks and implement masks and would be trained to do so. However, this is contingent on the portable oxygen tanks being immediately or being made quickly available. Thus, we assign a low likelihood score, 2 out of 5. The severity of not implementing oxygen masks is high, as the crew would be living in an oxygen-poor environment and that could lead to serious health concerns and loss of life. Thus, we assign a severity score of 4. This makes the criticality score that this implementation strategy would elicit the first generic safety control flaw a 8. Repeating this exercise for the remaining five generic safety control flaws, we obtain a criticality score for this implementation strategy of 80, or a medium criticality following the guidelines in Table 6.3.

Additionally, we consider the probabilities contained in the implementation strategy effectiveness. The first probability, the probability that the implementation strategy is implemented perfectly, we consider to be high. The crew will be trained to be able to successfully use oxygen masks, and astronaut training involves extensive underwater activities. In addition, we do not expect the crew to spend a significant amount of time finding and using the oxygen tanks if clearly labeled and in an appropriate storage space. However, this may be a possibility. Thus, we assign a probability of 0.8 that this implementation strategy is implemented perfectly. The second probability that we consider is the probability that this implementation strategy is available for use. Because portable oxygen tanks are a finite resource, it is possible that they may no longer be available when it comes time for use, having already been used up for many other possible reasons. Thus, this is a highly volatile probability and difficult to predict. But, since portable oxygen is a finite resource, we must account for the possibility that it may not be available, so we assign a probability of availability of 0.5. The third and final probability that we consider is whether the implementation strategy will be competent to control against the source, or the lack of oxygen. We know from experience that oxygen tanks are highly robust against low oxygen environments, and we can confidently say that

a crew member implementing an oxygen tank and mask will be able to operate adequately in a low oxygen environment. Barring a rupture of the tank or the mask slipping off, we consider the implementation to be competent and assign a probability of 0.9. Finally, the active control time of the implementation strategy is short, on the order of several minutes, and highly depends on the position of the crew member in relation to the position of the oxygen tank. We do not expect the crew member to take more than a couple of minutes to find and implement the oxygen tank, if available, and thus the active control time is on the order of minutes.

We repeat this process of assigning criticality scores, probabilities, and timescales for all the implementation strategies of the safety control option space for this hazardous state and create the risk decision matrix shown in Figure 6.3.

Hazardous State:	Oxygen concentration in habitat is too low										**Critical**				
Safety Control:	Ability to provide oxygen masks and containers to the crew					Ability to isolate oxygen supply to specific areas of the habitat					Ability to increase output of oxygen generation system				
Implementation 1:	Bring and store portable oxygen tanks that the crew can use to maintain normal oxygen levels and still be able to get around the habitat					Implement that at least one compartment in the habitat has the required minimum amount of oxygen concentration in the air, even if it means other rooms have insufficient oxygen					Implement a reserve oxygen generation system in addition to the nominal oxygen generation system				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	80	0.8	0.5	0.9	~ [m]	118	0.7	0.7	0.7	~ [m]	94	0.8	0.9	0.9	~ [m]
Implementation 2:	Implement reserve oxygen storage tanks and drop oxygen masks from the ceiling or walls for the crew to use					Modularize the habitat oxygen supply so that a failure in one compartment only affects the oxygen concentration in that compartment, and the crew can continue in the other rooms					Implement oxygen candles in addition to the nominal oxygen generation system				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	122	0.6	0.7	0.7	~ [s]	136	0.6	0.8	0.8	~ [s]	134	0.9	0.5	0.7	~ [m]
Implementation 3:	Implement that the crew wear their spacesuits until the oxygen levels have been restored					Implement multiple, physically separated habitats so that a failure of the oxygen generator in one habitat does not mean the other habitat(s) is unlivable					Provide more power to the oxygen generation system so that the chemical reaction to produce oxygen is increased				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	36	0.9	0.6	0.7	~ [m]	128	0.3	0.2	0.6	~ [hrs]	136	0.8	0.7	0.8	~ [s]

Figure 6.3: Risk Decision Matrix for *Low Oxygen Concentration* Disruption

Using Figure 6.3, we can actively select controls and implementations of those controls that are less likely to be flawed, have high probabilities of success, availability, and competence, and can be completed quickly. Control effectiveness considers all these measures, and at first glance we can select an implementation of a safety control that exhibits high, or green, performance in each of these categories. For example, we can further consider the first implementation of the third safety control, implementing a reserve oxygen generation system, because it exhibits high success probabilities and can be implemented quickly. Additionally, we may decide to no longer consider the third implementation strategy of the second safety control, implementing multiple habitats, because it may be very difficult to implement perfectly, may take a long time to implement, and it may be difficult to maintain two or more working habitats that are always available. We can

modify our decision-making process by implementing more stringent guidelines on the values in our control effectiveness metric, reflecting the risk averse perspective shown in Table 6.4. By doing this, we create the risk decision matrix in Figure 6.4.

Hazardous State:	Oxygen concentration in habitat is too low										**Critical**				
Safety Control:	Ability to provide oxygen masks and containers to the crew					Ability to isolate oxygen supply to specific areas of the habitat					Ability to increase output of oxygen generation system				
Implementation 1:	Bring and store portable oxygen tanks that the crew can use to maintain normal oxygen levels and still be able to get around the habitat					Implement that at least one compartment in the habitat has the required minimum amount of oxygen concentration in the air, even if it means other rooms have insufficient oxygen					Implement a reserve oxygen generation system in addition to the nominal oxygen generation system				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	80	0.8	0.5	0.9	~ [m]	118	0.7	0.7	0.7	~ [m]	94	0.8	0.9	0.9	~ [m]
Implementation 2:	Implement reserve oxygen storage tanks and drop oxygen masks from the ceiling or walls for the crew to use					Modularize the habitat oxygen supply so that a failure in one compartment only affects the oxygen concentration in that compartment, and the crew can continue in the other rooms					Implement oxygen candles in addition to the nominal oxygen generation system				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	122	0.6	0.7	0.7	~ [s]	136	0.6	0.8	0.8	~ [s]	134	0.9	0.5	0.7	~ [m]
Implementation 3:	Implement that the crew wear their spacesuits until the oxygen levels have been restored					Implement multiple, physically separated habitats so that a failure of the oxygen generator in one habitat does not mean the other habitat(s) is unlivable					Provide more power to the oxygen generation system so that the chemical reaction to produce oxygen is increased				
	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active	Criticality	P_perfect	P_available	P_competent	t_active
	36	0.9	0.6	0.7	~ [m]	128	0.3	0.2	0.6	~ [hrs]	136	0.8	0.7	0.8	~ [s]

Figure 6.4: "Risk Averse" Risk Decision Matrix for *Low Oxygen Concentration* Disruption

We use our control effectiveness metric to consider architectures and design choices that can provide the necessary level of resilience against these kinds of hazardous states and disruptions. Like most risk mitigation techniques, the considerations of probability and severity are necessarily subjective. To gain a more complete picture of the effectiveness of each safety control, as well as the risks associated with each safety control, this process could be done collaboratively with representatives from each sector of the design team. For example, to grade implementation strategies that mitigate risks associated with low oxygen environments, it may be necessary to include team members from air quality control, members that work on ECLSS, chief engineers and program managers, and system safety engineers. This is a popular practice when completing risk mitigation in industry, and this control effectiveness technique should take advantage of the collaborative culture involved in risk management. Additionally, control effectiveness is not meant to be an automated process in selecting the "most effective" safety controls. The vector of attributes in Equation 6.1 highlights different aspects of each safety control, which should be considered both individually and as a set. The values should not be combined in some way to yield a single value, as such a metric would hide the nuances of each safety control.

Finally, we use control effectiveness to provide design guidance from a system safety perspective. We are investigating design solutions and risk mitigation techniques that provide a certain level of resilience compared to other options we may consider. The design recommendations that we gather from this exercise aim to keep the habitat operating in a region of safe behavior, or to return the habitat from a region of unsafe behavior to a region of safe behavior. We do not consider other important design factors like cost, launch mass, technology readiness level, system performance measures, or human factors. To achieve a space habitat design that satisfies as many design constraints as possible, the recommendations produced by the control effectiveness metric should be considered in a design space in combination with these other important design factors in the interest of feasibility and cost, among others.

7. CONCLUSIONS

7.1 Summary

In Chapter 2, we introduced our control-theoretic approach to risk analysis. We explain how we derive our control-theoretic approach from the idea that we treat safety as a control problem. We define regions of safe and unsafe behavior for our system as nominal and hazardous states, and we design safety controls to prevent the propagation from nominal to hazardous state and to return the system to from a hazardous state to a nominal state. We also introduce the state and trigger model, which we use to model our sets of disruptions, hazardous states, accident states, and safety controls. We end the chapter by showing how our control-theoretic approach is one way of accomplishing the larger system safety process. We map our approach to the system safety process, and describe how we use our framework to identify disruptions and hazardous states, assess these disruptions and hazardous states, design safety controls to mitigate these disruptions and hazardous states, and assess the effectiveness of our safety controls.

In Chapter 3, we demonstrated how we used the first two steps of our approach to identifying safety controls to identify a preliminary list of disruptions and hazardous states for our example Martian habitat. This chapter covered the first step of our control-theoretic approach to risk analysis and is one way of completing the *Hazard Identification* step of the system safety engineering process. We also in Chapter 3 introduced the database of disruptions and hazardous states. We created a database in Microsoft Access to collect and store the relationships between the disruptions and hazardous states and used this database to visualize these relationships in a directed network format.

In Chapter 4, we showed how we can use traditional hazard assessment techniques like a Preliminary Hazard Analysis (PHA), System Hazard Analysis (SHA), and Operations & Support Hazard Analysis (O&SHA) in combination with our failure network to assess our disruptions and hazardous states and prioritize mitigation with our safety controls. Chapter 4 covered the second step in our control-theoretic approach, which is to assess disruptions and hazardous states. This chapter helped demonstrate how our process is compatible with the overall system safety engineering process and can be used in the *Hazard Analysis* step of risk management.

In Chapter 5, we used the third steps in our approach for identifying safety controls to demonstrate how we use safety controls to mitigate disruptions and hazardous states. Steps 3a and 3b in the approach for identifying safety controls are developing safety controls and generic safety controls. We used these safety controls and generic safety controls for an example habitat to inform the third step in our control-theoretic approach, using those safety controls to mitigate our disruptions and hazardous states. We developed a safety control option space for our example habitat and showed how we store safety controls developed using the state and trigger model in the database and in the network format. This chapter showed one method of completing the *Hazard Controls* step of the system safety engineering process.

In Chapter 6, we used the fourth steps in our approach for identifying safety controls to demonstrate how we assess the effectiveness of our safety controls in mitigating disruptions and hazardous states. We identified safety control flaws and generic safety control flaws and used these flaws to inform our control effectiveness metric. We used the control effectiveness metric to assess the effectiveness of our safety controls, and this covered the fourth step of our control-theoretic approach for resilient design. This chapter showed one method of completing the *Verification of Controls* step of the system safety engineering process and completed our mapping of our control-theoretic process to the traditional risk management process.

7.2 Key Findings

This thesis helped in validating our control-theoretic approach to resilient space habitat design. Our four step process of identifying disruptions and hazardous states, assessing these disruptions and hazardous states, using safety controls to mitigate these disruptions and hazardous states, and developing the control effectiveness metric to judge the effectiveness of our safety controls demonstrates that we can approach risk management from a controls perspective. Rather than assessing hazards, or individual faults that may or may not lead to hazards, we assess the system based on its state and whether it is operating in a safe or unsafe state. Our approach for identifying safety controls informs our control-theoretic approach, and iteratively adding disruptions, hazardous states, and safety controls based on generic sets of safety controls and safety control flaws allows for effective brainstorming of what can go wrong, what we can do about it, and how we can do it. Our control-theoretic process is grounded in system safety engineering and fits within

the traditional risk management framework; however, we use novel techniques to identify and assess hazards and mitigation techniques. Further, our control-theoretic approach to resilient design is based in modeling the system as a series of states and triggers between those states. Although this thesis describes a specific application of our approach, any system that can be modeled as a series of states and triggers can be a suitable application for our approach. For example, Landry (2010) modeled a system consisting of multiple aircraft using a state-based approach and used a set of triggers to and away from an end state *Loss of Separation.* Considering a set of aircraft as a system (or system-of-systems), we envision the application of our approach to use safety controls to mitigate the many disruptions and hazardous states present in the air traffic control network.

To assist in the development of resilient space habitat systems, we proposed a way to systematically develop a database of disruptions, hazardous states, and potential safety controls. We used a state and trigger model to model the habitat's states and transitions between these states, and a network model to illustrate the relationships between the disruptions, hazardous states, and safety controls. Using these relationships, we determined how a disruption could propagate through the habitat systems and trigger subsequent hazardous states, which we mitigate by implementing safety controls. We used our database, network format, and network measures to create an environment where we are able to assess disruptions, hazardous states, and safety controls individually, analyze the number of relationships and types of relationships between nodes of interest, and study from a top-down approach what disruptions, hazardous states, and safety controls may be the most critical to the habitat system. In combination with traditional hazard and mitigation assessment techniques, a failure network containing disruptions, hazardous states, and safety controls is a useful tool to investigate what may go wrong in a system and what mitigation strategies are available to address these hazards.

Finally, our control effectiveness metric is intended to indicate how well a control addresses the hazardous state or disruption that it was designed to control. We use our generic safety control flaws obtained in our approach for identifying safety controls to indicate how a safety control may elicit those flaws, and develop specific implementation strategies for our safety controls to investigate how a safety control may achieve its control function. Within these implementation strategies, we consider certain probabilities of success, availability, and competence in the face of

a source of a disruption or hazardous state, and we consider the time it may take to implement the control. Culminating this data on certain safety controls and their possible implementations, we can provide insight on how successful the safety control may be against controlling a disruption or hazardous state. This is one way of assessing mitigation techniques and can be used in combination with other control verification techniques to validate that the controls are adequate to control a hazardous state or disruption. Along with other considerations like system performance, cost, launch mass, and crew considerations, we can use control effectiveness to assist habitat designers in designing resilient space habitat architectures.

7.3 Limitations and Potential Improvements

Our exercise in identifying disruptions and hazardous states was limited to leveraging previous human spaceflight experience, extracting incidents and accidents, and extrapolating those possibilities to surface habitats and their unique extraterrestrial environments. Long term space habitats are complex and unprecedented systems and identifying all possible disruptions and hazardous states is a challenging task. We made various assumptions about what systems would be present in an example habitat, how these systems would interface, what kind of disruptions are present in an extraterrestrial environment, and how these disruptions would propagate into hazardous states. As the design of the habitat matures, more information should become available about the composition of the habitat systems, the habitat subsystems, and their interfaces. Completing this control-theoretic approach iteratively throughout the design process is imperative to capturing as many and as diverse a set of disruptions and hazardous states as possible.

While the RETHi team has identified a large set of disruptions, hazardous states, and potential safety controls, as well as relationships between them, the database and from there the network still needs to be expanded significantly. First, we expect that the set of disruptions, hazardous states, and potential safety controls will always be added to as we learn more about the habitat. Second, the database does not currently include the accident states that may ensue from hazardous states. Third, the network currently only includes a subset of the potential relationships between disruptions, hazardous states, and potential safety controls. The RETHi team is working to add transitions from one hazardous state to another, beyond the current transitions from subsystem to system to habitat. Further, safety controls themselves can cause transition to hazardous states given

the occurrence of certain flaws. Finally, we currently do not include transition probabilities—these could be integrated through edge weights, for example.

We presented an approach to assessing disruption and hazardous state criticality based on both the "local" probability and severity and the importance of these nodes within the network. We considered the sum of indegree and outdegree as a simple and intuitive metric of a node's importance. Additional measures, such as measures of centrality and path lengths, may be useful in assessing criticality as the more complex relationships mentioned in the previous paragraph are added.

An identified limitation of our process is that it involves a fair amount of subjectivity. When assessing the values involved in the control effectiveness metric, many come down to engineering judgment and leveraging of experience (traditional risk assessment methods face the same challenge). We attempted to avoid subjectivity when analyzing the relationships between the disruptions and hazardous states and which disruptions and hazardous states prioritized controlling, however both traditional hazard analysis techniques and our network format have elements of subjectivity. Identification of disruptions and hazardous states is dependent on the experience of the person doing the identifying, and our process for identification of safety controls is systematic but a brainstorm in nature. We use a systematic, in-depth grading scheme for control effectiveness, but inputs to the metric may vary depending on the person assessing the safety controls and implementation strategies. For potential improvement, we will use the RETH institute to test the control effectiveness and continually improve the metric throughout the length of the project. The RETH institute has demonstrated the ability to provide an end-to-end framework of modeling several scenarios for an extraterrestrial habitat, and we can use the computational and physical testbed in development by the institute to conduct trial-runs and improve the fidelity of the control effectiveness metric.

Lastly, system safety engineering is an inherently collaborative discipline. To adequately identify disruptions and hazardous states, assess those disruptions and hazardous states, develop safety controls and mitigation techniques, and assess those safety controls, it will be important to involve stakeholders that are not only experts in the many constituent systems of a space habitat but who also can provide input on the kinds of environments that extraterrestrial habitats may be exposed

to. For example, if a solar panel is covered by dust and the suggested safety control is to task a robot agent with cleaning the solar panel, that may require input from the power systems engineer, the robotics engineer, the engineer in charge of the health management system, a crew member may have to monitor the robot and monitor sensor readings, etc. This is just one safety control for one hazardous state, and it is not guaranteed that this safety control is the most effective control for this hazardous state. When evaluating safety controls, it is important to involve subject matter experts on each system involved as well as system safety engineers. This is common practice in most risk management exercises, and our proposed control-theoretic approach should be no different.

REFERENCES

Bahr, N. J. (2016). *System Safety Engineering and Risk Management.* Boca Raton: CRC Press, Taylor & Francis Group.

Dunbar, B. (2013, 05 09). *International Space Station*. Retrieved from National Aeronautics and Space Administration: https://www.nasa.gov/mission_pages/station/expeditions/expedition35/e35_050913.html

Frazier, S. (2017, 08 07). *Real Martians: How to Protect Astronauts from Space Radiation on Mars*. Retrieved from National Aeronautics and Space Administration: https://www.nasa.gov/feature/goddard/real-martians-how-to-protect-astronauts-from-space-radiation-on-mars

Jackson, S., & Ferris, T. (2012). Resilience Principles for Engineered Systems. *Systems Engineering*, Vol. 16, Issue 2.

Knight, J. C. (2002). Safety Critical Systems: Challenges and Directions. *International Conference on Software Engineering* (pp. 547-550). Orlando: ICSE.

Leveson, N. (2004). A New Accident Model for Engineering Safer Systems. *Safety Science*, Vol. 42, No. 4, pp. 237-270.

Leveson, N. (2011). Applying Systems Thinking to Analyze and Learn from Events. *Safety Science*, 55-64.

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems. *Organizational Studies*, 227-249.

Newman, M. E. (2003). The Structure and Function of Complex Networks. *SIAM Review*, Vol. 45, No. 2, pp. 167-256.

Panteli, M., & Mancarella, P. (2015). The Grid: Stronger, Bigger, Smarter?: Presenting a Conceptual Framework of Power System Resilience. *IEEE Power and Energy Magazine, vol. 13, no. 3*, 58-66.

Patriarca, R., Bergstrom, J., Di Gravio, G., & Constantino, F. (2018). Resilience engineering: Current status of the research and future challenges. *Safety Science*, Vol. 102, pp. 79-100.

Perera, J. S. (2012). NASA's Risk Management System. *Dutch National Risk Management Conference* (pp. 1-20). Ede, the Netherlands: NASA.

Rao, A., & Marais, K. (2020). A state-based approach to modeling general aviation accidents. *Reliability Engineering and System Safety*.

Rasmussen, J. (1997). Risk Management In a Dynamic Society: A Modelling Problem. *Safety Science Vol 27, No 2/3,* 183-213.

Rose, A., & Liao, S. (2005). Modeling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions. *Journal of Regional Science*, 45(1): 75-112.

Uday, P., & Marais, K. (2015). Designing resilient systems-of-systems: A survey of metrics, methods, and challenges. *Systems Engineering*, 18 (5), 491-510.