Finding the Beginning to Discover the End: Power System Protection as a Means to Find the First Principles of Cybersecurity

A Thesis
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Electrical Engineering
in the
College of Graduate Studies
University of Idaho
by
Nicholas C. Seeley

Major Professor: Brian K. Johnson, Ph.D.
Committee Members: Yacine Chakhchoukh, Ph.D.; Joseph D. Law, Ph.D.
Department Administrator: Joseph D. Law, Ph.D.

December 2021

Abstract

Engineers responsible for designing the protection, control, and operational technology (OT) systems used within electric power systems continue to adopt new technology even as some of this technology presents a clear and seemingly unfixable risk to the continued safe and reliable operation of the power system itself. These risks are directly related to the ability of malicious actors to compromise, control, and hold at-risk devices protecting and monitoring critical infrastructure. This thesis focuses on the "seemingly unfixable" aspect of the technology and how that can impact the operation of the electric power grid. In this thesis I explore developing the first principles of cybersecurity specific to OT. As first principles, they should apply more broadly to all of cybersecurity; however, this thesis focuses on operational technology. I use these first principles to take a look at simple power system protection architectures and describe why the cybersecurity problem is so deeply entrenched. From there, I explore how to use these notional first principles to derive a relative measure of system security. Finally, I conclude with recommendations for future research.

Acknowledgements

The University of Idaho ECE faculty that I encountered while taking classes over the past several years were all fantastic. In particular, Drs. Hess, Johnson, and Law all provided me with an education that was better than I could have hoped for. I feel it important to acknowledge that their passion for teaching and helping me to learn has played a not insignificant role in helping me get to where I am today professionally, and for that I am grateful.

Dedication

Given my less than stellar academic performance during my undergraduate education, I didn't think that any university would accept me into their graduate school. When I told my wife about this fear, she shook her head and said something along the lines of, "That's ridiculous." She was right.

Table of Contents

List of Figures

Chapter 1: Introduction

I work a lot with microprocessor-based protective relaying, and in my experience, cybersecurity has always been a consideration when designing these products. However, there is certainly a cost-benefit analysis in terms of both user experience and economics that becomes part of the decision-making process when designing relays. Protective relay designers ask: how secure is secure enough? How would we know? How would we know when we have reached the point of diminishing returns? What dictates these limits? What are the cybersecurity fundamentals that need to be considered when designing a protective relay? Are there first principles that govern the concept of security?

The consequences of ignoring these questions seem obvious. There are countless examples of how protective relays, or other industrial control system devices, are compromised, leading to results that range from unfortunate to truly terrifying [1] [2] [3] [4]. At present, my perception is that the most talked about methods to address these issues advocate for using additional technology. To me, this only seems to compound the problem and make it more intractable. How can the solution be more technology when this very technology enabled these problems in the first place?

When I first started to think about this problem, I began to get very concerned. How was it that an entire industry started and progressed to such an advanced stage without understanding the first principles behind the problems that have manifested? My concern only demonstrated my ignorance and lack of critical thinking towards the issue. This was made plainly obvious after I read an article by Landwehr, where he pointed out that it is not unheard of for technological revolutions to begin well before first principles are discovered [5]. Landwehr uses the example of the Duomo cathedral in Florence, Italy, which was built 350 years before Newton—at a time when the science of mechanics had not been formalized. Yet, the cathedral is still standing to this day. Is technology outpacing our ability to secure it in the cyber world? Yes. Is this to be expected? Using history as our guide, yes.

This search for the first principles of cybersecurity, or the “science of security,” as it is more popularly referred to, is not new. The National Security Agency has sponsored research in the area for a number of years now [6].

I am approaching the topic through the technology that I am most familiar with, power system protective relaying. As a result, the metaphors that I use and the analogies that I draw are common in the power engineering field but may not be as familiar to the typical cybersecurity readership. I do my best to make the concepts relevant and understandable to both fields, that said, I’m a power engineer, not a computer scientist.

1.1 First Principles

So, what is the root of the problem? I started by asking: what are the first principles of cybersecurity? Two years later, after reading hundreds of papers and numerous books along with hours of discussions, I am a little closer to an answer, but in true scientific exploration form, I generated more questions than answers. I thought the questions might be quickly answered, especially after the first paper I read had a specific section titled “Fundamental Principles” [7]. To my disappointment, the section largely dodged the idea of philosophical first principles and provided what I would call “prescriptive characteristics” of a secure system: need-to-know, including hardware, software, and physical security measures, etc. That revelation became a common theme throughout the papers and books that I read—the explanations of fundamentals and first principles did not rise to the level that I would consider as proper treatment of the concept of first principles; at best I would consider them functional characteristics[1].

So, I kept reading and enlisted the help of others that have made information security a significant portion of their careers. This group spent many lunches together over the last 24 months obsessing over small but vital details in word usage and definitions, having nuanced discussions of concepts like trust, complexity, and uncertainty.

The words that follow are my attempt to characterize the process that I went through to uncover the concept of first principles in cybersecurity: where I am, how I got there, and what I recommend as a path forward. Many helped in this effort, but what follows are my interpretations and conclusions drawn from my research and conversations with others. In terms of the usefulness of this document, I would like to think that it can provide a roadmap for similar future investigations or provide insight on how one goes about discovering first principles or thinks about problems in terms of first principles. I feel that such efforts, when done with purpose and intent, are valuable exercises in gaining true understanding of a given topic. In the end, I have combined the work of others, added my own insights, and produced a method of formalizing the relative cybersecurity of protection systems. I would expect that this revelation will be useful to the industry, if for no other reason than to spawn additional research to refine the concept and expand it to broader applications.

To the best that I have established, the first principles of cybersecurity are rooted in uncertainty, invocation of trust, and unavoidable complexity. I am not the first person to address the first principles of cybersecurity; however, my focus is more strongly influenced by philosophy and social science than other factors that I have found. Shouhuai Xu advocates for a framework of cybersecurity dynamics where the fundamental principles of cybersecurity are placed on a four-dimensional axis with first principles modeling, data analytics, and security metrics over time composing the four axes [8]. The bulk of the

[1] A term I borrowed from [22].

related work seems focused on establishing metrics, though brief philosophical discussions of the concept of security, trust, resilience, and agility as applied to cybersecurity are discussed [9]. I certainly appreciated the discussion of first-principle modeling in [8]; however, the actual discussion of first principles seemed more related to discussion of defense rather than the philosophical first principles that I had been searching for. That said, I will reference Xu extensively throughout this thesis, as his ideas have influenced this thesis greatly.

All this research has led me to think of several possibilities of why there does not already seem to be volumes of research dedicated to this specific topic. One, it is just not that important—related thoughts include: it is only an academic exercise that will yield little practical insight, and who cares? Two, it is abstract, and the results will not be realizable, practical, or practicable. Three, people have better things to do with their time than muse philosophically about cybersecurity. Finally, four, the demands of commercial markets have launched at a rate far outpacing security research. Regardless of the scenario, my musings still follow.

To summarize my findings, the best that I have discovered and can articulate would be stated in an overarching, broad-brush first principle as:

> *Cybersecurity has its origins in the notions of uncertainty and risk and is addressed through concepts of complexity and trust.*

Further, I will argue and formally prove a method of determining the relative security of cyber systems applied to power system protection.

1.2 First Principles as Applied to Protective Relaying and Power Systems

After discussing these low-level, fundamental topics, I will relate the discussion back to the concept of power system protection and protective relaying. Protective relays act to monitor, control, and protect the electric power system. This thesis assumes working knowledge of protective relays and their functions; hence, I will not be giving an overview of their functionality. Suffice it to say that when people in industry worry about the cybersecurity of the electric power system, they are, in a large part, worrying about the cybersecurity of protective relays and the communication networks that connect the protective relays.

Throughout this thesis, we will look at and understand how protective relays are designed and manufactured. This understanding will provide us the insight allowing us to draw conclusions about the susceptibility of protective relays to cyber intrusion.

While the result of a cyber-intrusion in a protective relay may be obvious, the impact to the grid may not be so obvious. With this in mind, it is important to understand that while intrusion into a protective relay through a cyber-attack is an obvious failure of the system, the consequences of the attack can vary immensely based on the device, its function, the vulnerability, and how the vulnerability is exploited. We will discuss these nuances in detail towards the end of this thesis.

1.3 Definitions

Before we go too far, it will be necessary to define a few terms that I will use regularly throughout this text.

Trust: My working definition is formed from James Coleman [10], which we will describe at length later in this text. At this point, I will define trust as a binary decision made by the trustor; one either trusts or does not.

Trustworthiness: The subjective position, judged by the trustor, that the trustee occupies on the spectrum that is used to base the decision to trust.

Cyber system: Any group of electronic components connected through purpose built electronic means of communication that can operate semi-autonomously but are ultimately designed, developed, controlled, and maintained by human operators.

Cybersecurity: The notion that a cyber system is monitored and protected from malicious actors through process, controls, or both.

Chapter 2: Finding the Beginning

2.1 What are First Principles?

This question very quickly lands in the realm of philosophy. Two definitions/explanations that I found seemed the most philosophically sound, one from Aristotle and one from Descartes. Not considering myself skilled in the art of philosophy, "seemed the most philosophically sound" is short for: I have heard of both Aristotle and Descartes and have at least a basic appreciation of their contributions to science and scientific methodology and, therefore, trust their insights.

In Posterior Analytics [11], Aristotle describes the idea of "induction" as "argument from the particular to the universal." I found Gasser-Wingate's explanation to be the most accessible when he argues that Aristotle's idea of induction is:

"…a form of cognitive development that begins with perception and progresses through a series of increasingly sophisticated states in which various universal concepts come to be formed in our souls." [12].

While not actionable in any helpful form, Aristotle's idea can help us in a couple respects. One, it enforces the idea that a first principle starts with a specific idea and that idea is formed to represent a universal truth. Two, it implies that the initial idea or perception is the origin and, therefore, has no further constitution. More colloquially: it is what it is. Wingate-Gasser supports this interpretation when he wrote:

"The first principles from which demonstrations begin are explanatory primitives. Since demonstrations explain their conclusions, these first principles cannot themselves be demonstrated." [12]

This is important to understand because, by this logic, we cannot be expected to demonstrate a first principle. We should be able to demonstrate how the first principle is used to solve another problem, but the first principle itself is the axiom, the statement made that has no direct proof, but everything else can be explained in terms of it.

For example, consider Maxwell's equations. They can be used to explain almost all electrical/electronic technology used today. But to explain: *why* does a changing magnetic field produces an electric field? Because of physics. That is about as good as we can get. We cannot demonstrate why the phenomena that are described by Maxwell's equations exist. The phenomena are a first principle. They are the origin.

Descartes builds on Aristotle's idea of a first principle being the origin—the most fundamental element of which cannot be further broken down. In "The Principles of Philosophy" Descartes lays out two conditions a "first cause" must exhibit in order to be considered a first principle:

> …they must be so clear and evident that the human mind, when it attentively considers them, cannot doubt their truth; in the second place, the knowledge of other things must be so dependent on them as that though the principles themselves may indeed be known apart from what depends on them, the latter cannot nevertheless be known apart from the former. [13]

Descartes presents the idea of doubt and the necessity of doubt in searching for truth. Because we are capable of doubt, we must be able to think, and if we can think, we must exist. This line of logic leads Descartes to his most notable claim, *"I think, therefore I am…"* While greatly simplified here, these concepts become Descartes first seven principles of human knowledge. [ibid]

For the purposes of identifying the first principles of cybersecurity, starting at deriving the epistemology of human knowledge is likely beginning further back than is practical. However, the fundamental question exists, "What is the most elemental aspect of cybersecurity?" If we can start to chase after the answer to this question, perhaps we can get closer to arriving at a first principle.

I will be using Descartes's idea of first principles as my reference. Nevertheless, before we start this journey, we should revisit the history of cybersecurity and the seminal papers that have shaped the thought and progress on cybersecurity over the last 50 years.

2.2 A Brief Review of Seminal Cybersecurity Papers

The widely cited grandfather of all cybersecurity papers is the RAND Report by Willis Ware from the RAND Corporation [7]. This paper was the result of a commission sponsored by the US government in 1967 and through the agency that would later become the present Defense Advanced Research Projects Agency (DARPA). As the US government was one of the first large adopters of computer systems, it became clear in the mid-1960s that computer security was going to become a topic of major importance. The RAND Report highlighted several issues of concern when considering the use of computer systems for handling US government classified data. Unfortunately, a least in my estimation, the report spent little time identifying the nature of the issue—four paragraphs out of a 100+ page report—but focused heavily on prescriptive method of solving the problem.

This is an easy criticism 50 years later, and it is not leveled to minimize the importance of the recommendation that the Commission provided: need-to-know privileges, roles-based access control, logging, etc. It is likely that all modern security controls could be traced back to the recommendations from this report. However, it is also fair to say that these controls just do not solve the problem. Fifty years later, cybersecurity is just as relevant and likely in more disarray than in the beginning. However, to say that these early researchers missed the opportunity is to underestimate the difficulty of the problem. Not to mention, how could these men and women possibly have known the extent to which computers would permeated our lives fifty years later.

Shortly after the RAND report was published, James Anderson published a two-volume report for the US Air Force [14]. The Anderson report picked up where the RAND report left off and delved further into solutions without much discussion of the root of the problem. The paper astutely points out that designing security after the fact has little chance of being effective and makes a case against the use of "tiger teams" to find vulnerabilities in systems to achieve proper security. The paper advocates for a "reference monitor" that acts as the arbiter of system access, allowing or not allowing access to various levels of classified information.

My first reaction to this solution was the obvious question: "Who or what, then, is going to monitor the monitor?" This line of questioning will become an important one when exploring the deep recesses of first principles. However, for this background discussion, all we need to know is that developing solutions dominated the early conversations regarding cybersecurity, and to the extent that I can see, very little time was spent on defining the true nature of the problem. In that respect, not much has changed.

Despite Anderson jumping into solutions before the problem was properly defined, he very poignantly summed up the reason computer security was inadequate at the time—and I would argue still inadequate today:

> A large part of the design problem **is attributable to the absence of models** as a medium for translating security requirements to technical specifications and as a source of acceptance criteria for evaluating the product. Without such models, system developers are forced to apply ad hoc security related techniques throughout the design and implementation of the system. This approach inevitably leads to exploitable flaws, and makes the security assessments necessary for certification virtually impossible. [ibid, emphasis mine].

He rightly points out that a lack of models hampers any ability to adequately design a requirement. Where I feel he goes astray is that he never gets to first principles to develop the model. Having never developed these first principles to inform the models, how could we possible solve the cybersecurity problem? I relate to Anderson in his comment about "ad hoc security related techniques." It feels to me that modern-day cybersecurity is largely a suite of ad hoc programs and devices, marketing buzzwords, and fancy advertisements. However, I believe that the concept of first principles and the dedicated, rigorous study of the problem to arrive at first principles can get us to where we need to be; this is not to dismiss the significant contributions of others in the field. There are many other smarter and better thinkers than I that have tried to solve this problem.

Additionally, on a slight aside, Anderson, as part of his report, detailed the cost of developing an infrastructure for the US DoD that would allow a single computer system to host all levels of classified and unclassified data useable by people with all levels of security clearance (no clearance to a Top Secret clearance). Anyone with current knowledge of how classified computer systems work will recognize that such a system still does not exist (almost 50 years later) nor is it likely to in the future. To me, this just highlights the enormity of the problem.

Around the same time that Anderson was thinking about a reference monitor solution, Bell and LaPadula were working at the MITRE Corporation and assigned the task of a mathematical formalization that guaranteed secure access control [15]. While the details of the mathematical proof of assured security is well defined in the paper, there is a glaring issue with proving that the subsequent system that implemented these formally defined controls. Bell and LaPadula acknowledged the obvious question head on: how can you guarantee that the machine rules, as implemented, are inviolable? This means that the formal mathematical system would need to be coded in a machine without error. Not only would the code for the formalism need to be flawless, everything involved in how the computer operated would also need to be guaranteed to be flawless. They provide no answer to this question, but just acknowledge it as an obvious question.

I found it amusing that David Bell, during an interview in 2012, recorded the history of his groundbreaking work, recalled saying when he first received his assignment at MITRE to study and develop a formal method of computer security, "That sounds pretty boring" [16]. It seems rare that people initially dismiss something as boring and then proceed to become a world-renowned expert and make major contributions in the advancement of that very thing. Amongst being a great and influential figure in computer security, I equally admire him for his intellectual curiosity and open mindedness.

A couple years later in 1975, at MIT, Saltzer and Schroeder published a paper "The Protection of Information in Computer Systems" which appears to be the first published notion of what is now referred to as the CIA triad [17]. In addition, they proposed design characteristics of secure computer systems that are still largely in use today. This paper may include the first reference of two-factor authentication and makes a vigorous case for the importance of simplicity—a topic that we will revisit later in this paper. In addition, Saltzer and Schroeder assert that the design for computer security should be open and not based on secrets. This statement should not be construed to imply that passwords or keys should not be kept secret, but that the design and implementation of the system should not demand secrecy. In present terminology, such a requirement of security depending on secrecy would be described as security by obscurity. This is type of security is widely regarded as lacking, as evidenced by the case that Saltzer and Schroeder present in their paper.

Many of the characteristics of secure cyber systems that were written about some 50 years ago are still in use today. There are two particularly glaring exceptions. In Saltzer and Schroeder make mention of "Fail-safe defaults." This can best be described as deny-by-default. The very first switched computer network designs allowed all traffic to flow and across networks and to any connected machine, and it was up to the machine to authenticate. Largely this practice is still in place today. Why? Likely because of market forces. Yost writes:

"The goals of greater efficiency and keeping overhead down tended to trump strong security for most firms in the late 1970s, 1980s, and in many cases, beyond." [18]

The other exception relates to Andersons' critique of ad hoc security measures. He writes:

"Unless security is designed into a system from its inception, there is little chance that it can be made secure by retrofit." [14]

The focus on secure by design never really caught on. As a result, it seems that devices and software were designed and placed in service, found to have vulnerabilities and patched, or worse yet, encapsulated by a separate device or software that mediated the first vulnerability, but very well could have introduced more vulnerabilities. The vicious cycle is obvious. The sales cycle is equally obvious.

The list of papers goes on.

2.3 Starting With the CIA Triad

After reading a number of papers and having a number of conversations about the fundamentals and first principles of cybersecurity, one topic is consistently mentioned as the basis of all cybersecurity: the CIA triad—Confidentiality, Integrity, and Availability. These concepts are widely regarded as the pillars of cybersecurity [19] [17]. In short:

Confidentiality: Ensuring that only those with the required and approved need have access to the information

Integrity: Ensuring the information sent is the information received

Availability: Ensuring that, when needed, the information is available

The above definitions may likely come under attack for a variety of reasons, both reasonable and not. This is precisely why I began to question the idea of treating the triad as first principles: no one seems to agree on precise definitions. The ideas Saltzer originally defined are slightly but consequentially modified from how standards bodies like ISO define them. Not having a precise definition or statement makes it difficult, if not impossible, to satisfy Descartes' first condition of a first principle.

For example, ISO 27000, a family of international standards for managing information security risks and the associated controls to do so, defines confidentiality as a:

"property that information is not made available or disclosed to unauthorized individuals, entities, or processes" [20].

Compare this to Saltzer and Schroeder's original definition of failing to preserve confidentiality:

> An unauthorized person is able to read and take advantage of information stored in the computer. This category of concern sometimes extends to "traffic analysis," in which the intruder observes only the patterns of information use and from those patterns can infer some information content. It also includes unauthorized use of a proprietary program [17]

The difference in language is nuanced but striking. The obvious difference between the two is that Saltzer and Schroeder frame the definition around a perpetrator and protection from malicious acts, whereas the ISO 27000 definition frames the idea in terms of appropriately making information available. To me, this is the difference between protecting the information from malicious actors versus allowing access to those that need the information. The difference in mindset might be explained by the intent of the authors, where Saltzer and Schroeder, and most every cybersecurity researcher of that era, were working on projects associated with the US Department of Defense (DoD). The ISO organization, from Switzerland no less, takes a more neutral tone.

None of this discussion is to suggest that the language associated with the triad is responsible for the failure to make cyber systems implicitly safe. That said, I would be surprised to hear anyone make an argument that failure to standardize on language has helped the situation.

Insofar as language is concerned, different definitions will dictate one's ability to decide whether or not a security control is appropriately addressing the elements of the triad. Take for example the arguments made in [21]. Lundgren uses the example of a time-controlled safe and that unavailability is a method of making something security. According to the ISO definition of availability— "property of being accessible and usable on demand by an authorized entity"—a time-controlled safe does not meet the criteria. On the other hand, Saltzer and Schroeder's definition of compromised availability is fully compatible:

"…an intruder can prevent an authorized user from referring to or modifying information, even though the intruder may not be able to refer to or modify the information" [17]

The security behind a time-controlled safe is a perfectly acceptable control—depending on the needs of the user of the document—as it allows modification at a certain time of day but is locked at all others.

Similarly, and further complicating matters when applying the triad, at what point does a system become cybersecure? Is perfection demanded at all levels of the triad? As Lundgren argues, this does not seem to be the case. Additionally, if the triad itself is malleable and perhaps context dependent, it is hard to argue that it is a first principle, or even a fundamental characteristic of security.

To imply that the CIA triad is therefore unnecessary or unusable would be a misstatement. I believe it is useful in practice as a practical exercise to evaluate possible controls and also useful in helping us get to the true first principles. I would consider the triad as a functional characteristic, and in doing so, it was helpful to think about why. What is it about confidentiality, integrity, and availability that make them important aspects of cybersecurity? What are they rooted in?

These thoughts lead me to carefully consider Descartes' requirements of a first principle. Interpreting his words, I have found a first principle when the idea can no longer be broken into smaller, more fundamental ideas. So, what are the fundamental components of confidentiality, integrity, and availability?

This exercise led me to consider a topic that I am much more familiar with, applications of protective relaying in electric power systems. The first book that I ever read about protective relaying was GE's *Art and Science of Protective Relaying* [22]. The book lays out the functional characteristics of protective relaying as sensitivity, selectivity, and speed. I found there to be a striking resemblance between these three characteristics and the triad. In both, it seems that perfection of all elements is not possible or even necessary. Furthermore, improving one characteristic seems to negatively impact the others. With that, what are the roots of sensitivity, selectivity, and speed? Each characteristic for microprocessor-based protective relays has, at its root, at least one or some combination of Maxwell's equations, the Nyquist criterion, and Shannon's theorem.

For example, speed is dictated by the propagation of the electromagnetic wave produced by a fault on a line; speed is dictated by physics and defined by Maxwell's equations. As defined by the physics we know today, a protective relay will never be able to detect a fault and operate faster than it takes for the fault-induced wave to travel down the line where the relay sits. Maxwell's equations are the most basic elements that define the limits of speed in protective relaying; they embody the first principle of speed. Similar arguments can be made for why Shannon and Nyquist play a role. Can we use a similar process starting with the CIA triad to get to first principles?

2.4 A Foray into Defining First Principles

If confidentiality, integrity, and availability are functional characteristics of cybersecurity, what is their foundation? Starting with confidentiality, what is the genesis of confidence? If I tell someone to keep something in confidence, I am relying on my trust in that person to obey my request. If that trust is misplaced and the person betrays me, then my confidence is broken. I would argue that, similarly, if information is to be held confidential the underlying principle that allows that confidence is trust.

In the case of Anderson's reference monitor, confidentiality is preserved by the reference monitor overseeing interactions and requests for data. If the person or program requesting specific data is authorized to access the data, as determined by the set security policy and arbitrated by the reference monitor as the ultimate authority, access to the data is granted. Otherwise, access to the data is revoked. But as asked earlier, what is monitoring the reference monitor? If we perpetually need a monitor to monitor the monitor, we enter a never-ending cycle. At some point, we must trust that, in this example, the reference monitor is doing what it is designed to do.

On the other hand, we can consider Bell and Lapadula's formal methods of determining security. They used general system theory to derive and prove a well-defined mathematical system for guaranteeing the security (in this case, also confidentiality) of a system. The caveat being, in their words:

"Two problems are immediately evident. First, unless the system guarantees the inviolability of rule W our security theorem does not apply…" [15]

The actual rule W is inconsequential for our discussion; however, what is important is that the inviolability of W would be determined by the hardware and software that implemented the rule W. How do we guarantee the inviolability of hardware and software? My argument is that we cannot. At some point, we must invoke trust.

To dig even further, how do we guarantee that Bell and Lapadula's rule W is correct? They assert that the rigor with which they used system theory to develop the proof guarantees the viability of their assertion. But how do we know this is true? The main finding behind Kurt Godel's "On Formally Undecidable Propositions of Principia Mathematica and Related Systems" [23] which is beyond the scope of this work, could be used to argue that blind faith in the power of mathematical rigor may not be fully justified. At that, we are trusting that the mathematical foundation upon which Bell and Lapadula developed their proof is sound.

The same can be said for the characteristics of integrity and availability. How do we guarantee integrity or availability without, at some point, trusting the very mechanism that is used to ensure integrity or availability? It seems difficult to refute the idea that eventually we simply need to trust that what was designed as a security measure was designed without vulnerabilities and that it will function as intended. This, as best as I can tell, is a bitter pill to swallow because it is seemingly rarely mentioned in literature. Further, popular marketing seems to deny the role of trust altogether by insisting that “Zero Trust” is the gold standard for security [24].

As uncomfortable as it may sound, it seems that all matters of security come back to trust in some way, shape, or form. I say uncomfortable because how can we ever prove that a system is secure when a notion like trust is at the root? How do you, or even can you, quantify trust? Discussions on the concept of trust quickly depart into the realms of sociology and philosophy because of the abstract nature of the topic. It would be hard to imagine that anyone over the age of 10 does not understand, at some level, the idea and basic premise of trust. However, for such a familiar concept, it turns out that it is incredibly difficult to understand the “why” of trust, and even the “how” of trust is difficult to pin down.

However, before we get into the why of trust, it is important to ask: is trust the first principle? Going back to Descartes, we again ask if trust is the most fundamental element, or can it be broken down further? With that question, it is helpful to ask another, why does trust exist in the first place? This question leads into volumes of texts and theories on the origin and nature of trust that people have spent lifetimes dedicating themselves to the study of. This is to say that there is no one universally agreed upon explanation, but what I found most convincing was from James Coleman and his book *The Foundations of Social Theory* [10]. Coleman argues that issues of trust are a subset of issues that result from risk, and without risk, trust has no need to exist. Following this line of thought, if trust is a subordinate to risk, this stands to reason that risk is more fundamental than trust.

In a similar exercise, we can ask, is risk the most elemental concept? To this question, we might need to figure out why risk is present. We can make the argument that risk would be non-existent if all data and results were certain. From the perspective of cybersecurity, if one knew with absolute certainty that the only ones that had access to, and would ever get access to a critical system, were those who had been appropriately authorized, then all malicious outsider security problems go away. With absolute certainty, risk seems to disappear because what is risk other than a calculation to quantify the likelihood that an event goes as expected. This would explain why risk is often measured in probability.

So, what is the first principle of cybersecurity? Uncertainty. The root of the problem of cybersecurity, as far as I can tell, lies at the feet of uncertainty. How useful is this revelation? Not very, in a practical sense. It only allows us to ask more questions but does not seem to provide any real solution. However, uncertainty, at least to me, is more promising as a first principle than trust in terms of holding hope for an analytical solution to the problems of cybersecurity.

Uncertainty is a topic of considerable study in modern physics and impacts our knowledge of all physical systems. Proposed by Werner Heisenberg, what has become known as the Heisenberg Uncertainty Principle, is a highly regarded, yet not proven, hypothesis regarding the unknowability of "conjugate pairs" of information [25]. We will dive into this in greater detail in later chapters, but it is suffice to say, uncertainty may provide a physical basis for defining the limits of security.

Since the term uncertainty by itself does not present much to build off of when unearthing first principles, we need to begin the difficult work of succinctly phrasing the rationale for why uncertainty is the first principle. We start with the first principle and follow by structuring subsequent arguments that develop a framework to convey something useful and actionable.

I posit the first several principles of cybersecurity as follows:

1. Complete knowledge of a system is unobtainable; therefore, uncertainty will always exist in our understanding of that system

To secure a system means to protect it from harm or malintent. Protecting a system from harm means to completely understand what composes it. To completely understand what composes the system, one needs complete knowledge of all components that compose the whole.

2. The principal of a system must invest trust in one or more agents

Without certainty, we are required to invoke a mechanism to address the gap in true understanding of the system. Given that cyber systems involve sharing and dissemination of information, they have characteristics of a social system. Social systems often employ trust as a mechanism to mitigate uncertainty. Situations that require trust are situations that involve risk via uncertainty. The total risk associated with trusted agents in a system comprises both known and unknown risks.

3. Known risks can be mitigated using controls, transference, and avoidance, else the risks must be accepted

While expressly rejecting the idea of complete knowledge of a system, elements of a system can be known. In the case of cybersecurity, anything known is known through experience. Experience informs mitigation techniques. This experience allows us to devise controls. This experience also allows us to realize that controls may not be effective, and effectively transfer the risk to another part of the system, or different system altogether, or we can choose to avoid the risk in a variety of ways (e.g., not implement the system that presents the risk). If any of the three techniques are absent, we are tacitly accepting the risk.

4. Unknown risks manifest through complexity

Complexity coupled with uncertainty creates a fog which diminishes our ability to discover conditions that lead to security risks. Without complexity, risk in cybersecurity is a function of uncertainty alone. As uncertainty is addressed by trust, all risks become known as they align with where we placed trust. Complexity, on the other hand, obscures the relationship between the system and where trust was placed within that system.

The four principles listed above present a few concepts that we will need to explore in greater detail in the following pages. In particular, uncertainty, trust, and complexity are well known concepts, but there may be considerable disagreement in a universal definition for each. The preciseness of language and awareness of semantics becomes crucial when we try to define these terms and discuss how and why they are critical concepts in cybersecurity. As noted above, Descartes' framework for establishing first principles requires the argument be, to paraphrase, so "clear and evident" that their truth cannot be doubted; hence, careful attention to words and language is of dire importance.

As I write the statements above, I understand those reading will fall somewhere on the spectrum of agreement to disagreement. The goal in the following chapters is to provide the exposition and convincing arguments to defend and advance the first principles described above. The end goal of establishing the first principles is to develop a framework that will help us define at least some methods of how to best provide security to cyber systems. Ultimately, I believe these first principles will enable us to determine and define a fundamental limit of cybersecurity, though it is not the intent of this thesis to get us all the way there.

Chapter 3: What is Uncertainty?

Uncertainty is often times associated with probability. I can only imagine this is done to provide some semblance of comfort to the idea that we understand a problem well enough to make a useful decision or interpret results. If we look at this through the lens of cybersecurity, I argue that probabilities play a neutral and, in many cases, detrimental role depending on how they are used. Is there any use to claim that this system has a 60% chance of being attacked, or this program has a 40% chance of having a vulnerability? Likely not, and for several reasons.

Luckily, we don't hear much about security professionals making such probabilistic claims, and if we do, the claim is likely, "this system has 100% change of being attacked," or "there is 100% likelihood that this program has a vulnerability." But such claims are largely not useful because they provide no actionable information. However, there is still a claim of certainty, or near certainty, in the above two statements. So, how do we come to this certain conclusion, and can these conclusions be made useful?

Probabilities are based largely on assumptions. In order for the probability to have any value, the assumptions must have value, and this is where we get into problems with cybersecurity. The foundation of our systems that are being protected and the cybersecurity systems that are implemented as the protector are so complicated with so many unknowns, assumptions about the composition of these systems are not much more than guesses. However, this gets us closer to understanding what uncertainty is.

Uncertainty, in the context of deeper philosophical discussions, is primarily discussed in relation to knowledge [26]. In the context of information theory, uncertainty points toward a lack of information [27]. This information theoretic approach to uncertainty provides important insights to our understanding of cybersecurity. As we will explore in this chapter, uncertainty can be thought of as a lack of information.

3.1 Information Theory and Shannon Entropy as an Exploration of Uncertainty

One cannot research the foundations of information theory by looking for the works of past researchers that have drawn a parallel between information and entropy, and to that extent, the parallel between information theory and thermodynamics [27] [28] [29]. The general idea is that as entropy increases, information is lost. In the context of a thermodynamic process, consider the melting of ice. When water is frozen, it assumes a rigid crystal lattice structure, and when the ice melts, the resulting water is less structured. We can say that we lost information as the ice turned to water. Without knowing specifics, we could generally say that we have information about the structure of the water as ice that we no longer have with the water in liquid form. The ice melted, entropy increased, and information was lost.

In the context of our discussion on cybersecurity, information is the actionable result of the data that we input into an algorithm. Using the example of a protective relay, information could be the decision to open a breaker or not. Where there is uncertainty to the question "does this circuit breaker need to open?", the algorithm takes in relevant data and outputs the information to answer the question.

From this simple example, we can start to formalize how we define the term "information." We will define information as the net change in the entropy of the system [29]. We will express uncertainty as equivalent to entropy and use the familiar equations for entropy. Hence, I will follow others and define uncertainty as [29]:

$$U = -K * \log P,$$

where *K* is a constant and *P* is the probability of an event. The constant *K* is arbitrary. In the formalism of thermodynamics where *U* is entropy, *K* is the Boltzmann constant with units of Joules per Kelvin. For now, we will leave *K* as unitless until we better describe exactly what information is in terms of units.

Given this, I will follow [30] and define information as:

$$I = U_1 - U_2 = -K * \log \frac{P_1}{P_2}$$

The simple point of these definitions is to formalize information as the change in uncertainty. If an event has a probability P_1 at a point in time and is then assessed to have probability P_2 at a different point in time, the change in the probabilities is directly proportional to the information added to, or subtracted from, the system. In other words, if the probability increased it is because information was added to the system to drive improved odds, and vice versa.

I'll take a moment to point out that information, as used here, represents knowledge directly applicable to the task at hand. We are all likely familiar with the phrase "information overload." The phrase implies that there is too much "information" to efficiently use during the decision-making process. I argue that the "information" referred to in the phrase information overload represents a combination of noise and information, and not truly the information that is relevant to the decision. The task that we have at hand is to understand what the true information is and differentiate that from the noise. We will talk later about the impact of this noise and how it drives complexity.

Following the example in [27], we can demonstrate the uncertainty and information calculation. Assume that we have two choices with equal probability. Sometime later, we are given information that narrows the decision down to the correct choice, i.e., probability of 1. We can calculate the information as follows:

$$I = -K * \log\frac{P_1}{P_2} = -K * \log\frac{\frac{1}{2}}{1} \approx 0.3 * K$$

Ignoring what K actually means, we see that there is a net gain in information *I*, i.e., *I* is positive.

As simple as this example is, it very much helps to frame the notion that the goal of cybersecurity should be to reduce uncertainty in a system. Reducing the uncertainty is the same as increasing the information within the system. The mechanisms for reducing uncertainty in any system using first principles are discussed in the following chapters. These chapters focus on the concepts of trust and complexity. As we begin to develop heuristics for understanding and controlling trust and complexity within a cyber system, we should be able to prove their benefits to the security of the system.

Chapter 4: The Elements of Trust

For the sake of clarity, focus, and relative brevity, we must restrict a discussion of trust to the elements of trust that are germane to our discussion. While an important end goal would be to quantify these elements in a way that allows us to formalize how trust is used in systems, the actual formalization is beyond the scope of this thesis. Instead, we will discuss these elements of trust, why these elements of trust are important to cybersecurity, and review prior work addressing formal methods of trust. The goal of this discussion is to develop an understanding of what trust is and how it relates to and impacts our issues of cybersecurity.

Of particular influence on my thoughts regarding trust are the works of Luhmann and Coleman. Even though their work focuses on applications relative to the social sciences, it seems more applicable to security in ways that the work of others, who focused on philosophical arguments of trust, does not. These less applicable arguments are based on conditions that appear to me as solely human characteristics, such as optimism [31], cooperation [32], and hope [33]. While cyber-related systems are largely created and used by humans—where optimism, cooperation, and hope might be applicable—they also involve machine-to-machine communication. This communication, I will argue, must still exhibit elements of trust. As such, imputing human concepts on non-human subjects leads to suspect arguments. Given the task of arguing that trust—which may already be considered a purely human concept—is necessary between non-human components, I will avoid adding any more fodder for the critics than necessary.

I will start by using Coleman's description of trust, as condensed by Harwood [34]:

1. *The trustor cannot achieve his ends, or cannot achieve his ends economically, without the collaboration of the trustee.*

2. *The trustee commits to actions on behalf of, or in the interest of, the trustor (let us call these the trustee's obligations), but the trustor has no control over the trustee's actions and so is subject to uncertainty or risk.*

3. *If the trustee fulfills her obligations, then the trustor will be better off than if he had not trusted the trustee. If the trustee fails to fulfill her obligations, the trustor will be worse off than if he had not trusted the trustee.*

4. *Generally, the trustor places resources at the disposal of the trustee but has no control over what the trustee does with these resources. So, the trustee may use those resources for her own benefit, for the benefit of the trustor, or for the benefit of both.*

5. *As a result of the trust, the trustee is in a position to do something that she could not otherwise do.*

6. *Finally, generally, there is a time lag between the commitment to the transaction by the trustor and the fulfilment by the trustee.*

While not the one or two sentence definition that I would prefer to work with, Coleman's description of trust seems general enough to accommodate our needs when discussing cybersecurity. The statements above are obviously aimed at and applicable to human social interaction, but below, I will explain their applicability to non-sentient systems. I will use the relationship between an electric power generator and a generator protective relay whose purpose is to protect the generator in question. In this case, the trustor is the generator and the trustee is the protective relay. Addressing Coleman's description of trust in the order set above:

1. The generator's ends are to continually and reliably produce electricity. The generator is not able to economically produce power long term without a mechanism to identify and protect it from hazardous situations—we will refer to those situations as faults.

2. The protective relay commits to identify faults and protect the generator from catastrophic damage caused by faults. The generator has no direct or indirect control of the protective relay.

3. If the protective relay detects a fault and protects the generator from damage, the generator is saved irreparable damage. Otherwise, the generator is likely to be subject to a catastrophic, unrecoverable fault.

4. The generator places resources at the disposal of the protective relay by way of current and voltage measurements. These measurements are the means by which the protective relay will determine if there is a fault present in or near the generator. Though, as Coleman indicates, the protective relay is not obligated to use the measurements, nor can the generator force the relay to use the measurements

5. Because of the protective relay, the generator is able to operate at full capacity, whereas without the protective relay, it might be recommended that the generator run at a greatly reduced capacity to limit the damage from a fault.

6. There may be a significant time lapse-conceivably years-between the time the protective relay is applied and the time a fault occurs on the system that requires action by the protective relay.

It would be understandable to criticize the idea that the generator is consciously trusting the protective relay as would be the case when fitting a human-to-human, or even a human-to-machine, interaction to Coleman's criteria. That said, the reason I find Coleman's description of trust so relevant is exactly because it doesn't seem to require a human element. The example above works and meets Coleman's criteria without either the trustor or trustee being sentient, provided that we are open to the idea of trust being applied to non-sentient objects in the first place.[2]

Having established a mechanism linking the machine-to-machine interactions to trust, and assuming that reader understands how Coleman's criteria applies to human-to-human and human-to-machine situations, we will discuss how trust impacts a cyber system. My examples will relate to industrial control system equipment. Even if the reader is unfamiliar with the exact purpose and function of the systems described below, I believe the context will be obvious enough for anyone familiar with basic internet-connected or networked devices that are commonly used in academic, office, and even personal environments.

In "Trust and Power" [35], Niklas Luhmann identifies several elements that are essential to trust. While I did not find all arguments to be of consequence to this discussion of cybersecurity, I will present the ones I found most compelling.

4.1 Trust and Time

Luhmann starts by acknowledging that trust has a direct relationship with time.

"It needs no more than a cursory inspection to show that the theme of trust involves a problematic relationship with time." [ibid]

The easiest way to understand how this relationship with time impacts trust in cybersecurity is to examine cryptography. We do not need a detailed understanding of how cryptography works to see that cryptography as a method of security (confidentiality in the CIA triad) is time limited. In general, the better the method of cryptography, the longer it will take to decode an encrypted message, but it is still a finite amount of time. Compounding this time limitation of trust is not just the theoretical security of the cryptographic algorithm but the practical implementation of that algorithm. It is not uncommon to find algorithms that contain unknown vulnerabilities, or have been implemented lazily and exploited long before a numerical cracking of the encryption has been accomplished [36].

Further, technical controls are not the only mechanism that are subject to deterioration over time. Even the most trustworthy of people can become less trustworthy through mental illness such as dementia, schizophrenia, etc. The reliability of trust placed in an operator of a cybersecurity system suffering from some mental ailment may be suspect. This is especially tricky if the person was perfectly mentally fit at the time he or she was first trusted with the position but then began to degrade over time. Similarity, part

[2] If one is not open to this idea, then the following arguments will likely ring hollow. However, those having difficulty accepting the idea of trust between non-sentient object may want to focus on the idea that the cybersecurity problem is ultimately a human to human problem, regardless of the number of non-sentient objects separating the business or social interactions occurring in "cyber space."

of the reason that candidates undergoing security clearance background investigations must disclose mental health issues is related to the trust that must be placed in an individual to safeguard national secrets [37].

While the relationship seems obvious enough, the period of time over which trust degrades is far from clear. Luhmann argues that trust is only valid in the present.

> Trust can only be secured and maintained on the present. Neither the uncertain future nor even the past can arouse trust, since what has been does not eliminate the possibility of the future discovery of another past. [35]

While insightful, this is hardly helpful in a practical sense because we are constantly in the present, but the security control was implemented in the past. The most we can take from this is to reinforce the idea that the trust placed in the security control at the moment the control was "present" cannot be assumed to be the same trust now that the security control is "past."

How quickly does it degrade? I would describe the answer to that question as unknowable, though we may be able to generalize relative time. The trust degradation due to the ailing mental faculties of an aging security professional is likely to be on the order of tens of years, whereas the trust degradation of a security control such as cryptography is likely to be in years. Certainly, there are defining moments that can force a step change in our trust. On the technical side, quantum computing is regularly heralded as the modern-day cryptography killer [38]. It may be wise to assume that degradation of trust at the hands of the technology could be based on known technical capabilities and hypothesized future technologies. To me, it seems that most technologies are not stumbled upon by complete accident, nor are they fit for use immediately upon discovery. This gives us at least a starting point for predicting when future technologies may supplant existing technologies, while the only certainty is that our prediction will likely be wrong.

Time, then, is of critical importance to trust and, by extension, cybersecurity. Any attempt at developing a formal method for describing and extending trust must include time as a key role in that development.

4.2 Trust and Familiarity

Luhmann writes that familiarity is a condition of trust in that one cannot trust something that that is completely unknown. Familiarity allows for past experiences to be incorporated in the decision to trust. I freely admit that this argument may seem difficult to justify for machine-to-machine interactions where the concept of familiarity is difficult to impute on non-sentient machines. However, I will argue that the concept of hashing algorithms and machine learning based statistical analysis types of controls can be reasonably described as mechanisms to promote familiarity.

Hashing algorithms are commonly applied to data as an integrity check so that a device receiving the data can check that the data has not altered in route. The basic idea is that the data-generating device uses an algorithm to generate a numerical representation for the data it is sending, the hash. The end device receives the data and the hash. It then computes its version of the hash based on the same algorithm that the sending device received and compares it to the hash that it received. Assuming the data sent is the same as the data received, the hashes should be identical. If the two hashes are the same, the receiving device declares the data valid and proceeds to process the data as intended. If the hash is different, the data is normally discarded and flagged as bad.

I argue that hashing algorithms act as a primitive form of familiarity. The receiving device is using an algorithm common to the system. Each pair of devices exchanging data are using this familiar algorithm to determine the integrity of the data received. The whole idea of integrity is to check the wholeness or soundness of the thing in question. This very concept implies that there must be a reference with which to check—in our case—the data. In order for this to be the case, all devices must be familiar with that reference. In this case, the reference is the hashing algorithm.

More sophisticated applications of trust as a function of familiarity are related to using communication traffic analysis to determine cyber intrusion [39]. While the techniques vary, the basic idea seems to use datasets that establish a basis of normal, typical network traffic behavior. With the knowledge of the typical network traffic, abnormal traffic patterns may indicate a compromise of the system. Similar to the argument above, understanding abnormal based on comparison[3] with normal requires familiarity with the normal.

Considering these examples, I find it reasonable to argue that familiarity is applicable to machine-to-machine communications as well as person-to-person. Accordingly, quantifying trust through familiarity, while not easy, is an essential part of understanding the impact of trust on cybersecurity. Trust is more well founded when interacting elements have a higher level of familiarity. Likewise, distrust can also be more well founded given more familiarity. However, unlike time, familiarity cannot be easily measured. This makes familiarity more difficult to consider when attempting to formalize trust.

4.3 Trust as a Reduction of Complexity

The title of this section is the same as the title of a chapter in Luhmann [35]. It is the simplest way of describing the relationship between trust and complexity. Trust allows us to reduce the complexity of our present and future, at least locally. I do argue that the complexity of the overall system is maintained, and

[3] It is fair to object that these more sophisticated detection algorithms are not operating solely on the basis of comparing. That said, a detailed account of how these algorithms work is unnecessary for this discussion and beyond the scope of this paper, so I will continue to use my overly simplistic language.

trust does not reduce complexity on an absolute scale; however, the complexity as presented to the security operator is greatly reduced. There are many examples of this, so I will only explain a single one that I think nicely illustrates the argument.

Consider cryptography. How many people actually know how cryptography actually works? Having not done any formal of scientific polling, I can only make a generalized argument based on observations working in a technical field, but I would argue that, as a percent of the number of users of the internet, the number is well under one percent. I base the number as a percentage of internet users for a specific reason. One of the most common security tips that internet security "experts" recommend is to ensure that a user only does financial transactions through websites using https protocol—https denoting that the session information is encrypted. Why would people use a technology for something as important as encrypting financial transactions without fully understanding that technology? The technology is complicated. It is based on some pretty gritty mathematical principles. It is not impossible to understand but certainly has a high barrier to entry. It is much easier to trust that the encryption works. Trust reduces the complexity to the user when engaging in online transactions. That said, the complexity still exists. The mathematics behind cryptography is still rumbling merrily along as the unwitting user goes about their everyday business. Local complexity, the complexity the user experiences, is reduced, but the system complexity still remains.

Without trust and its function of complexity reduction, it is hard to imagine that social and technological progress would have gotten very far. If we all had to understand the principles of lift and aeronautical engineering before ever getting on an airplane, the popularity of air travel would be greatly reduced. The same could likely be said about every other technology, not to mention every physical force. Would anybody be able to venture outside given our relative lack of understanding of gravity? We trust that it works the way we are familiar with and the way that people (scientists) claim that it works.

In terms of cyber-security, Luhmann can be invoked when he states:

"By means of trust, the trustor unburdens himself of complexity which he cannot sustain. Anyone who wishes to abuse his trust must take this complexity upon himself." [35]

This is an interesting lens to view cybersecurity through. Those that seek to compromise a system must overcome the complexity of the security control in order to defeat it. This is likely not what Luhmann had in mind when he wrote this, but it is seemingly relevant. Can we use this information to help us understand how to address complexity of a security control? At that, perhaps complexity is not an enemy of security. We will discuss complexity later in this thesis. It is suffice to say, complexity is a challenging topic in which to make progress, but at least it has more quantifiable properties than familiarity.

4.4 How Trust is Employed

I have just argued that trust, as related to cybersecurity, is predominately a function of time, familiarity, and is invoked to reduce of complexity. Exactly where does this trust lie when we consider cybersecurity system? We will use the example of a protective relay and map out all of the major components of the protective relay and identify where we would expect trust to play a role.

Figure 4.1 shows the architecture of an SEL-501 relay. At a high level, the 501 is designed to detect an increase of current flowing through a power line and issue a command to open a circuit breaker to stop the flow of current if it passes a user-programmed threshold. The architecture of Figure 4.1 shows the electronic devices that are used to create that device. We do not need a detailed description of how each device works, but as we get further into the discussion, we will focus on a few of these components and discuss their function when needed.

As alluded to earlier, the SEL-501 is designed to protect a single power line from an overcurrent situation that would be potentially damaging to the line itself or equipment upstream of the line. It accomplishes its design objective by opening (i.e., tripping) a circuit breaker when it detects an overcurrent condition. If the relay is compromised, an attacker would be able to use the circuit breaker tripping functionality of the relay to maliciously turn power off to downstream consumers fed from that particular line. In order to avoid unnecessary disruption of service, great care must be placed in protecting the device from malicious actors. We will focus specifically on the ability of the relay to open a circuit breaker.

The arrows in Figure 4.1 point from the component that is the trustor to the component that is the trustee in the trust relationship. The output contacts—labelled XOUTx or YOUTx in Figure 4.1—are responsible for the actual opening of the breaker. When given a command to open the breaker, the output contacts operate to open the breaker. In Figure 4.1, we see that the output contacts are trusting the microprocessor to make this decision. The output contact does not decide on its own to open the breaker; it receives a command from the microprocessor. The microprocessor, in turn, is trusting information that it is getting from other components, like the EEPROMs, SRAM chips, or the UART that is controlling external communications to the relay.

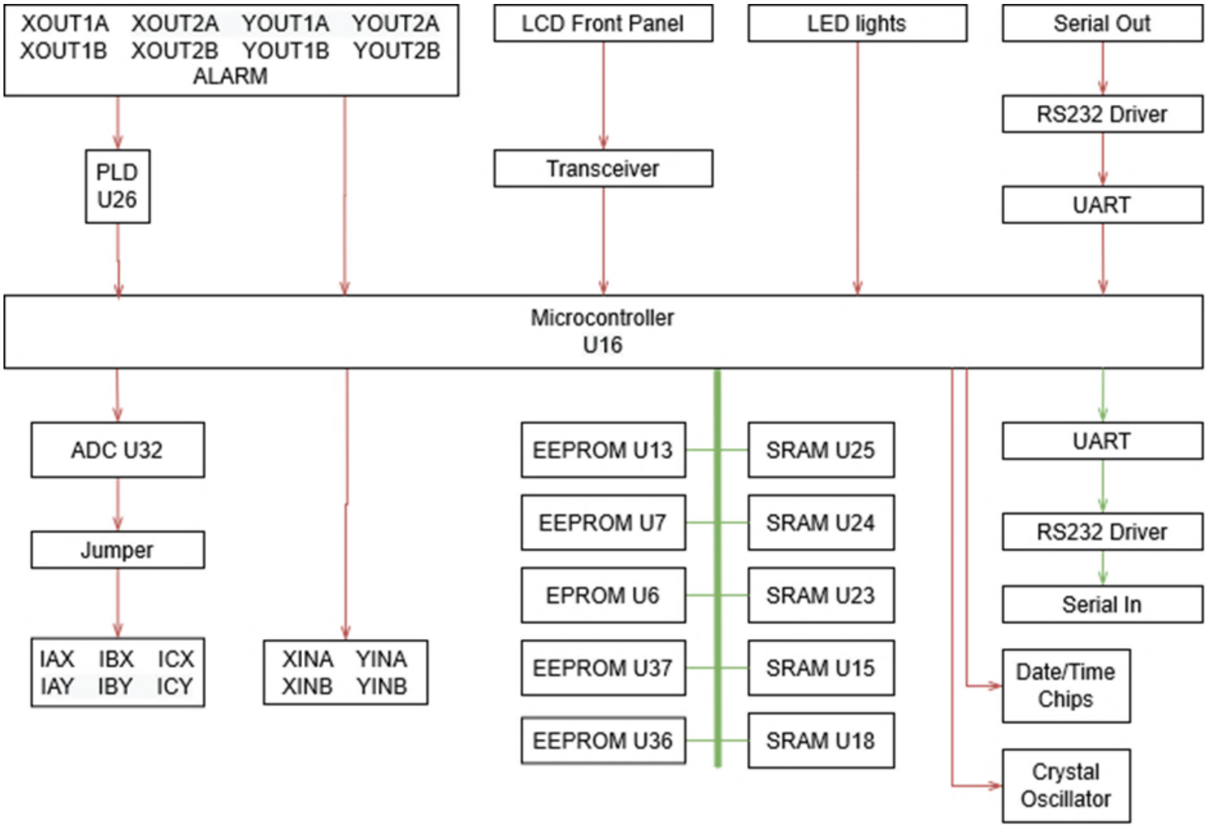

Figure 4.1 High-level topology of an SEL-501 Relay[4] illustrating trust relationships

At a higher level, the idea of trust becomes even more evident. A user is trusting that the device will function as designed and intended without knowing or understanding the bulk of how the relay works at the electronics level. This situation clearly exemplifies Luhmann's argument that trust is a mechanism to reduce local complexity, and if we think of why the user might be willing to extent this trust, familiarity likely plays into the equation. Those that will use the SEL-501 relay are likely either familiar with it or with the manufacturer. Through some means, the user applying this relay to their electric power system will require familiarity with the product.

To explore and understand the idea of familiarity we can ask, exactly how familiar can the user be with the product, and how far does that familiarity reach? Trusting a product on the basis of familiarity has some interesting challenges. If the user is familiar with the company and extends trust to the company and/or the product, the user is thereby tacitly trusting that the company manufactured the device

[4] A big thanks to Aron Lum for creating this diagram

according to secure practices. Beyond that the user is also trusting the manufacturer's supply chain, all of the companies included in that supply chain, and so on. This can become a circular chain of trust.

Evidence of thorough supply chain security practices is quickly becoming essential for companies supplying equipment to critical infrastructure [40]. I would argue that this is due to an increased need to validate the trust the companies involved in critical infrastructure are placing in their vendors. Stories about supply chain compromise are becoming more commonplace and the topic has even attracted the attention of the president of the United States, as evidenced by Executive Orders 13920 [41] and 14028 [42].

Trust appears to be unavoidable. The complexity of the relays, their manufacturing, and the relationship between all parties involved with the production of the relays, is intractable. Any attempt to quantify and validate the entire chain of the design, engineering, manufacturing, and application process is overwhelming. If this is not evident, we can expand Figure 4.1 to abstract the individual components of the device but include all elements that are involved in the production of that device.

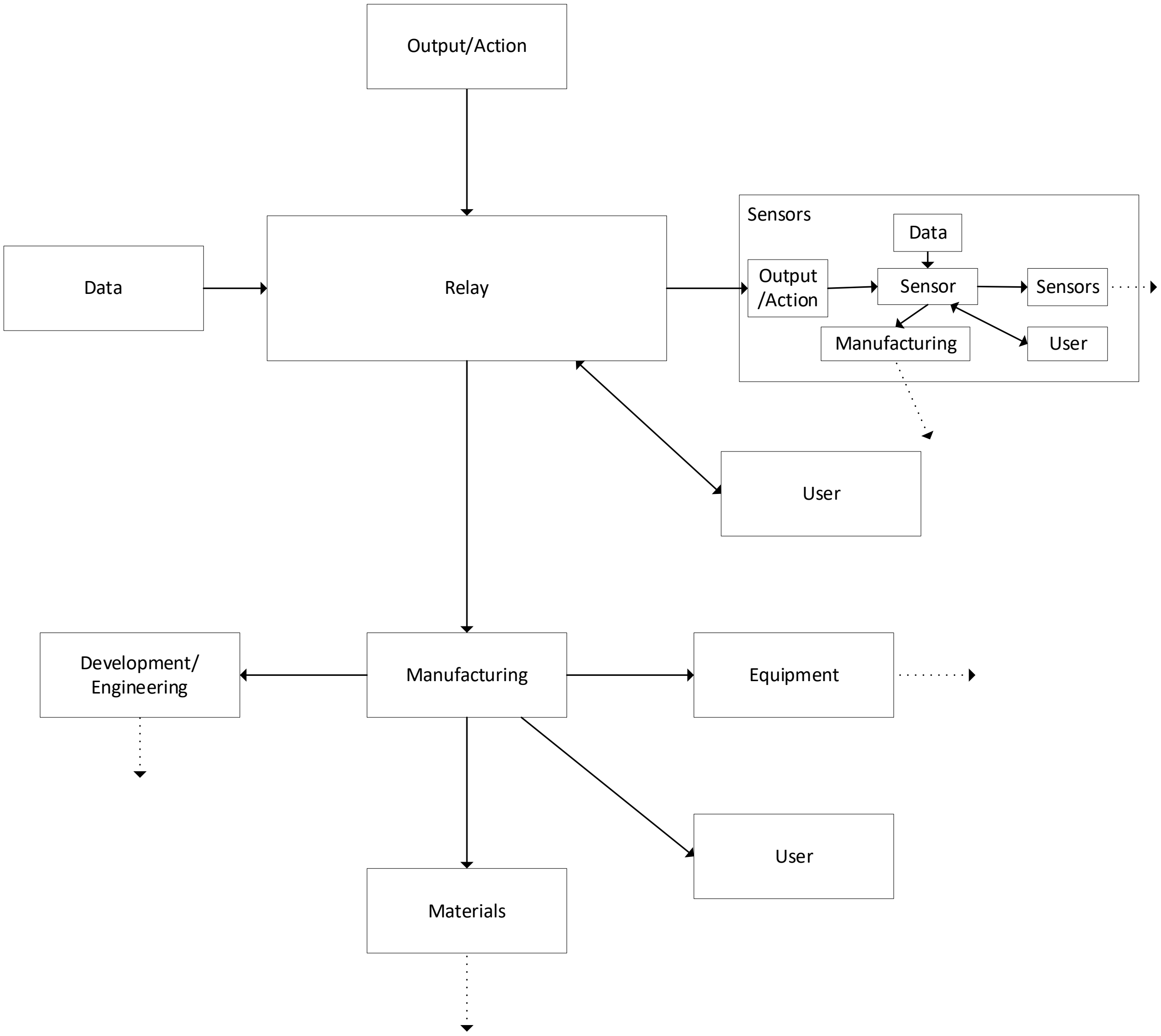

Figure 4.2 High-level view of where trust is placed in a system

Again, as in Figure 4.1 the arrows in Figure 4.2 point to the trustee. We can see, even in this high-level abstraction, a pattern emerging. The topic of complexity will explore these types of patterns in more detail. What is important to understand is that almost every box in Figure 4.2 can be expanded out using a similar pattern. It is likely that every box within a box in Figure 4.2 can also be expanded, and so on. The degree of nesting is extensive, and likely can circle back on itself, creating a never-ending loop of dependencies.

In light of this, how far back must familiarity extend to create or sustain trust? Ultimately, this depends on who is extending the trust, but practically speaking, there is some evidence in present supply chain security requirements being implemented by electric power utilities in North America as a result of NERC CIP-13 compliance requirements. The North American Transmission Forum (NATF) supply chain security framework does recommend that vendors to critical infrastructure industries document country of origin of their components [40]. This request directly extends beyond the boundary of the supplier and into the boundaries of the suppliers' suppliers.

4.5 A Brief Review of Formal Methods of Trust

Coleman derives a very simple equation for trust in his book *Foundations of Social Theory* [10].

$$Decision\ to\ Trust\colon \frac{p}{1-p} > \frac{L}{G}$$

$$Decision\ to\ not\ Trust\colon \frac{p}{1-p} < \frac{L}{G}$$

$$Indifferent\colon \frac{p}{1-p} = \frac{L}{G},$$

where:

p = chance of receiving gain (probability that the trustee is trustworthy)

L = potential loss if trustee is untrustworthy

G = potential for gain if the trustee is trustworthy

While the equations are simple and straightforward, they involve a tremendous amount of subjective judgement. It is likely that the most difficult variable to judge is the probability that the trustee is trustworthy, *p*. Knowing this variable to any certainty would be tremendously valuable. However, the crux of the entire issue of trust stands mostly on this judgement. Estimating the potential loss and gain are a little more obvious and calculable. However, I would suggest that unintended consequences will have a large impact in the total gain or loss. That said, unintended consequences have more to do with the situation with which trust is being applied and less so with the actual decision to trust.

Addressing the difficulty in determining the variable *p*, we need to realize that this variable represents the entire experience the trustor has with the trustee. These experiences will combine to allow the trustor to come to an assessment of the probability that the trustee is trustworthy. How the trustor will use these experiences to develop the probability of trustworthiness is something that we will discuss later. But first, we will turn to Marsh's work on formalizing trust to get a better sense of the role that time and memory play in our decisions to trust.

Marsh divides trust into three different categories: basic trust, general trust, and situational trust. Situational trust, he argues, is the most relevant form of trust when considering situations of cooperation [43]. While we are not precisely worried about cooperation in our situation, Marsh's concept of situational trust does seem the most appropriate for our uses. To formalize the concept of trust, Marsh presents an equation where situational trust is the product of importance, utility, and general trust.

$$T_x(y, \alpha) = U_x(\alpha) \times I_x(\alpha) \times \widehat{T_x(y)}$$

Descriptively, this equation defines the situational trust that x has in y for situation α $[T_x(y, \alpha)]$ as the product of the amount of utility x gains from situation α [$U_x(\alpha)$], the importance that x places in the outcome of situation α [$I_x(\alpha)$], and the assessment of the general trust the x has in y based on past experience [$\widehat{T_x(y)}$]. It should be noted that the result of this calculation will not produce a binary result, as would be consistent with my working definition of trust. Marsh goes on to use the results of this equation to then determine the decision to "cooperate" and this decision is binary. In short, I still stand by my assertion that trust is binary in that we either decide to trust, or do not, and it is trustworthiness that accounts for the spectrum that Marsh calls situational trust.

Marsh references Simon [44] when he describes the importance of utility [$U_x(\alpha)$] where the typical economic concept of a rational actor is used to justify the idea of maximizing utility. Because a rational actor is assumed to be interested in maximizing utility when making decisions, utility then becomes an important part of the decision to trust. Unless the trustor stands to gain from the decision to trust, there is little reason to extend trust. The more the trustor has to gain from trust, the more she benefits from the trustee upholding the trust; thus, utility is directly proportional to situational trust.

According to Marsh, importance [$I_x(\alpha)$] is "an agent centered or subjective judgement of a situation" [45]. Importance is used to accommodate the relative change in the impact that trust will have on a decision based on the context of the situation. If x considers trusting y for situation a, the importance of situation a is a key factor. Consider the implication of importance when accepting a bank check as a form of payment. If the check clearing the bank is the difference between being able to feed your family or not, the importance of the decision in accepting the check as a decision to trust is much greater than if the check will just go into a savings account, with no real immediate direct impact to the trustor. Hence, importance can change even if the actual decision is the same.

General trust [$\widehat{T_x(y)}$], as defined by Marsh is:

"It is x's estimate after taking into account all possible relevant data with respect to T x(y, α) values in the past"

This is Marsh's attempt to incorporate the history of interactions that the trustor has had with the trustee. This concept is similar to Coleman's variable *p*, the probability that the trustee is trustworthy. Where

Coleman's variable *p* should include all previous encounters with the trustee, Marsh uses a specific variable to aggregate the trustor's total experience with the trustee.

Note that time is not explicitly used in either of these formulations. As such, I would describe both of these formalisms as insufficient, though I have not done the work of offering an alternative of my own. Even though I would call them insufficient, it does not mean that they are not useful. In this respect, they at least attempt to give a user some concepts to consider when deciding how to measure trustworthiness to aid the decision to trust.

4.6 How Effectively Can These Methods Apply to Cybersecurity?

It would be relatively trivial to code Coleman's or Marsh's equations into a cybersecurity system, assign values to variables, and implement a trust-based cybersecurity algorithm. However, I would argue that the effectiveness of such an algorithm would be poor. The notions of *p* and *T* from Coleman and Marsh, respectively, are deep concepts that consider a tremendous amount of history. To effectively quantify these numbers would potentially be a large undertaking. Any effort to implement Coleman's or Marsh's equations would need to carefully consider how these variables are calculated.

Further complicating matters is that situations of cyber intrusion will likely be relatively rare. People have the relative advantage of learning to exercise trust over time, with numerous examples of both broken trust and upheld trust. This is most evident in the number of times the average person lies in a day [45]. People deal with lies that they are told every day of their lives through which they develop experience with how to judge people based on their actual and perceived honesty. This gives people the experience to learn from mistakes and re-evaluate their own criteria for trust (*p* or *T*).

In cyber systems, deceit is rare; therefore, learning how to detect deceit through experience is likely to take a long time, with several examples of broken trust. The consequences of trust are far graver when that trust is broken in these systems.

I would argue that without significant modification, perceived trustworthiness or general trust, as may be applied by people, is ill suited to the application of cybersecurity. While I still argue that trust is a valid concept for application within cybersecurity, the mechanism by which we develop trust for use in cybersecurity applications needs substantially more research.

Chapter 5: The Impact of Complexity

The inclusion of the concept of complexity in any discussion of a problem verges on a tacit admission that the prospect of solving the problem is hopeless. It is difficult to address the problem of cybersecurity without acknowledging the truly complex relationships involved in evaluating how secure a cyber system is. At a high level, cybersecurity is not achieved by a specific device that enforces all aspect of security but through multiple devices acting in concert to address all manner of security concerns that are present. These devices are designed, developed, manufactured, installed, and configured mostly by people. These people, like all people, have biases, finite mental acuity, and myriad other factors that complicate their ability to produce error free—in concept, intent, and execution—products.

Unlike the complex systems of biology or thermodynamics that seem to emerge spontaneously, the emergence of complexity in cyber systems may not be immediately recognizable. We've created these systems using our own knowledge, wisdom, and best practices. They did not occur spontaneously but pragmatically,[5] which is about the exact opposite of spontaneity. It goes to figure the complexity manifests itself through various means, and what makes a thermodynamic system complex is not necessarily what makes a cyber system complex. We will discuss several ideas behind complexity in the following pages.

5.1 Characteristics of complexity

Complexity takes many forms [46] but complex systems appear to exhibit common traits regardless of their specific natures [47]. There may be some question as to whether or not a cyber system would qualify as a complex system, but I will argue that including the entire system—both human and machine components—a cyber system certainly qualifies as a complex system.

Ladyman et al. sought to better understand complex systems and, in doing so, asked several scientists skilled in the art their definition of a complex system [48]. The responses had several commonalities; among them were hierarchy, emergence, and information conveyance, three characteristics that will be described in more detail below.

Revisiting Figure 4.2 High-level view of where trust is placed in a system, we see that many of the systems that compose a cyber system have similar constitution. This constitution appears to repeat,

[5] "Pragmatically" may be an oversimplification. As anyone that has developed commercial products would attest, economic and market pressures can greatly impact the design decisions made during product development, and "pragmatic" can give way to cheaper, quicker, and/or easier. Though relatively speaking, as compared to the emergence of a spontaneous process cyber system development appears fairly pragmatic.

perhaps to the point of creating a pattern, and apparent patterns, as Shalizi and Crutchfield argue in [49], seem to be an indicator of complexity.

Seth Lloyd created a list that categorizes various measures of complexity under using three questions: how hard it is to describe, how hard is it to create, and what is its degree of organization [46]. Thinking about complexity in these terms helps to sort out and concentrate on exactly how we should be thinking about complexity in the cybersecurity realm. One of the more common methods of analyzing complexity of a cyber system is to think of it in terms of complexity of description. In that, we can use the tools of information theory to help us understand how easy it is to describe a system.

In terms of describing complexity using concepts from information theory, some of the more common measures of complexity will be described below, but for now, we can think of a simple string of numbers. Say this system is a sequence of 1,000 numbers. If the sequence of numbers are all ones, the representation of that system does not need to be 1,000 ones long; it can instead be reduced to some notation that defined a one repeated 1,000 times, for instance, an algorithm that maintains the information contained in the string but reduces the size of the string required to express that information. However, if the string of numbers was completely random, there would be no way to reduce the size of the string because there would be no discernable pattern to the sequence by which to reduce its size algorithmically. In this case, I would argue the system is also of minimal complexity—perhaps even zero complexity—because there is no pattern. The idea of complex system implies there is a system in place, and anything identifiable as a system will have at least some minimally identifiable pattern. If the system has no pattern and is completely random, the idea of complexity doesn't seem to make any sense.

With this perspective, we can identify a relationship between the structure of information about the system and the complexity of the system. Below is a representation of the relationship. Note that the plot Figure 5.1 shows is a hypothetical representation of the complexity-randomness relationship, but there is no guarantee what the actual relationship looks like. Literature does not seem to converge on the shape that describes the actual relationship. Some representations show what looks like a normal distribution, others show a more piece-wise linear representation.

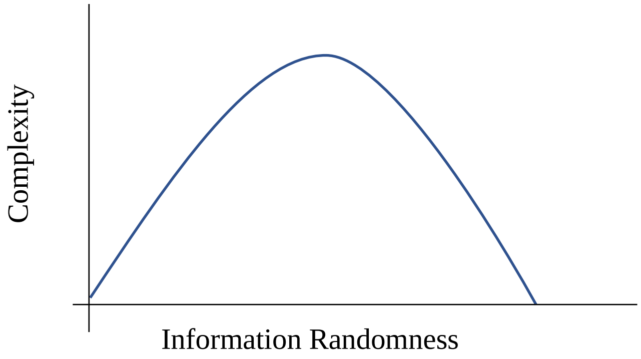

Figure 5.1 Complexity randomness relationship

While the shape of the relationship may not be well understood, there is wide agreement that complexity is minimal to zero at the extremes and peaks somewhere in the in between.

My guess is that, while we like to see symmetry in nature, perhaps there is no such symmetry in the relationship between complexity and randomness. If that is the case, what would Figure 5.1 look like?

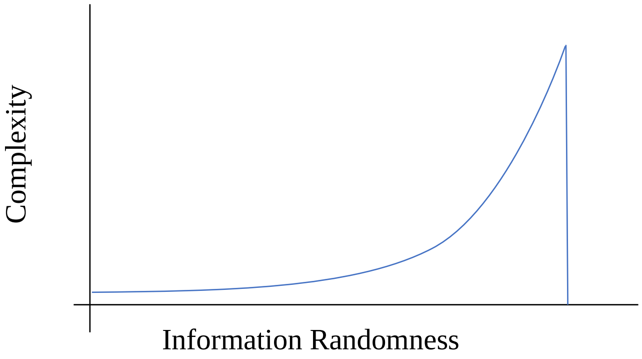

Figure 5.2 Non-symmetrical complexity randomness relationship

To me, it seems reasonable the complexity would increase as randomness increase until the exact point at which the system became completely random, at which point the complexity of the system would experience a discontinuity and exhibit zero complexity. Or, perhaps, complete randomness represents an asymptote of the system where a system will only ever get more complex but never actually reaches complete randomness, and therefore complexity only increases.

While researching this claim, I began to look for examples of system that began to decrease in complexity as randomness increased as illustrated in figure 5.1. I have not been able to find anything that convinces myself that any such systems exist, except where trust is applied, and even in the case of applying trust, total system complexity does not decrease, but local complexity certainly can.

While the above paragraphs make the point that complexity is a multi-faceted topic, we will start to explore the common characteristics that different complex systems can exhibit. Below, I explore two such characteristics. While they may not be universally accepted as fundamental characteristics of complexity, there has been enough written about these subjects in discussion of complexity that makes them defensible ideas.

Hierarchy

Herbert Simon wrote a widely cited paper on the topic of complexity in 1962 [50]. He defines complexity broadly, saying that a complex system is made of many components that interact in a *"nonsimple way."* I find this definition to be sufficiently broad to encompass a large number of systems and most certainly describe information systems and cyber-related systems. Simon argues that a foundation of complexity is hierarchy.

> Thus, the central theme that runs through my remarks is that complexity frequently takes the form of hierarchy, and that hierarchic systems have some common properties that are independent of their specific content. Hierarchy, I shall argue, is one of the central structural schemes that the architect of complexity uses. [50]

Simon goes on to say that the systems that exhibit complexity do so because they have had time to evolve; therefore, it is evolution that drives complexity. Each further increase in complexity is a product of starting at the last "stable intermediate state" and progressing to solve the next problem to attain a new stable intermediate state. These intermediate states drive the hierarchy of the overall system. Simon notes that the problem solving involved in getting from one state to another is a process of "…*selective* trial and error" [emphasis his]. This is an interesting notion in the context of cybersecurity that we will explore in more detail.

This explanation of evolution driving complexity is further elaborated in more modern interpretations of the relationship between complexity and hierarchy. Jessica Flack has a great example in [51], where she argues that:

> "…complexity and the multiscale structure [hierarchy] of biological systems are the predictable outcome of evolutionary dynamics driven by uncertainty minimization"

If we are to believe these assertions about the genesis of complexity owing, in part, to the formation of hierarchical systems, then we are certainly in good company. The impact of hierarchical systems is studied in many fields in the social sciences and well as biology, economics, and other hard sciences. To think of cybersecurity in terms of hierarchical systems is not intuitive, but once we open that line of thought and investigate the possibility, cybersecurity as a hierarchical system can make sense.

Cyber systems have been evolving since the mid-1960s and, in terms of technological evolution, have had plenty of time to evolve, especially given the time constant involving technology appears to be several orders of magnitude shorter than biological evolution. Each new technology gain that has been made over the last several decades has contributed to furthering the stable intermediate states cybersecurity has evolved to what it is now. Cryptography is a prime example of the selective trial and error process. Some of the earlier hashing algorithms have now been deprecated and are now no longer recommended for use [52]. Subsequent algorithms learned from the failures of prior algorithms to evolve beyond the present state to a new stable state.

On a system scale, advancement of the hierarchy within cybersecurity is no less obvious. When cyber systems were first created in the 1960s, they largely comprised a few computers used by researcher. Cybersecurity now impacts all forms of modern life, from communication to banking to research of all types. Nothing that we use regularly in our day-today lives can ignore the implications of cybersecurity.

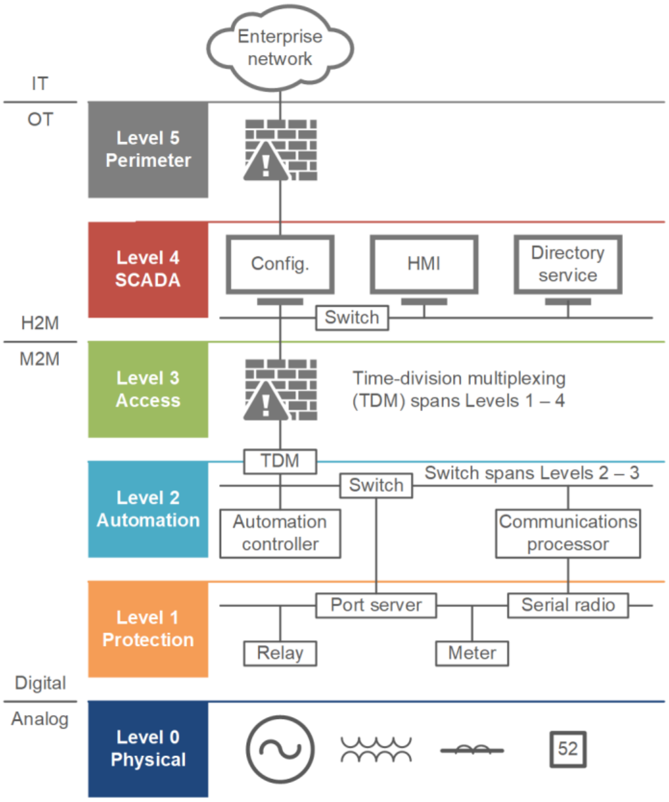

Figure 5.3 Layered defense model for ICS systems

As an illustration (Figure 5.3) of cybersecurity as a hierarchical system, we can look at the layered security network model that is routinely held up as cybersecurity best practice for industrial control systems [53]. This model describes the network as a system of systems and doesn't explicitly identify the even more complicating factors such as the complexity inherent to the devices themselves. The model is hierarchical in that the top level has control and omniscience over the lower levels. As one drills down to the lower levels the functionality is more specific and less aware of the upper levels.

In terms of cybersecurity in industrial control systems, I believe I have made a clear case that hierarchy is certainly a characteristic of the system. In doing so, I feel it is appropriate to maintain that a cybersecurity system is a complex system, partly due to its clear display of hierarchy. Another feature of a complex system, emergence, requires a little more nuanced argument.

Emergence

Simon also explores the concept of emergence as characteristic of complexity. The concept of emergence is widely applied throughout philosophical literature and writing dealing with complexity science; it is generally described as the whole being greater than the sum of its parts. On the other hand, in terms of a more philosophical framing of the property, emergence is demonstrated through a system that possesses both dependence and autonomy (or distinctness) when considering the role of the constituent parts [54].

Standish presented a more sophisticated way of explaining emergence by using the concepts of micro- and macrodescriptions [55]. Where microdescriptions are those that deal with the most fundamental level of a process and macrodescriptions are applicable to higher level process interactions, emergence appears when the characteristics defined using a macrodescriptions are not applicable within the microdescription.

If we apply these ideas of emergence to the premise of cyber systems and cybersecurity, we can gain some insight that helps us frame the problem. My initial thought centered around the idea that emergence cannot be shown using modern examples of cybersecurity devices on their own, only in the systems that they compose. I reasoned that, as complicated as these electronic devices may appear to an observer not skilled in the art, the parts are not greater than the whole. Instead, the device represents the obvious embodiment of the collection of the parts, which is to say that when you put a microprocessor in the device, one should reasonably expect that it will function with some limited, known problem solving capability. However, reading a report by Armstrong et al. convinced me to think more deeply.

Armstrong et al. make an interesting case:

> …[a] complex system has emergent behavior that cannot be predicted ahead of 'running' it, or alternatively, simulating that system with sufficient fidelity (and complexity) to see the emergent behavior [27] (Section 3). Concretely, this is why cyber systems that are composed of elements like programs, processors, and routers, each of which we presumably understand, are nonetheless constantly surprising us with the consequences of previously unknown vulnerabilities. [56]

Coming from a Newtonian physics background, it is hard for me to grapple with the idea that all information about a system could be known, yet the system can interact in ways that are unpredictable.

However, the idea of emergence in complexity means just that unintended functionality or capability emerges from the combining of multiple known functions [57] [58]. With this in mind, Armstrong and others are making the claim that vulnerabilities emerge from the interaction between the functions within a device. If this were true, it would seem that properties other than vulnerabilities would emerge as a result. It is possible that this is the case, but because people are only worried about and/or exploiting vulnerabilities, no one bothers to look at the other instances of emergence in these devices.

With this in mind, we should be able to make arguments for emergence at both the system level and device level. Starting with the system level, take for example a hypothetical cybersecurity system, as shown in Figure 5.3. Beginning in the early 1970s, the only part of Figure 5.3 that existed was the physical level and the protection level, and at level 1, the relay and meter existed in the 1970s. However, as time progressed and technology evolved, it became possible for relays and meters to digitize information and store that information, as well as receive control commands electronically via specific communication protocols. These communications and protocols became so central to the efficient operation of an ICS system that it became necessary to protect the systems from unintended and malicious acts. This has led to the creation of the present security community that is focused on inventing and establishing best engineering practice to properly secure a system. This security community is an ecosystem in and of itself, and it paved the way for new professions and new industries.

This is all to point out that as technology evolved, a completely new system emerged that, I argue, meets the very definition of emergence that was discussed above. Protective relays and meters evolved to include digital signal processing and memory. From that evolutionary step emerged an entire industry that dramatically altered the course of the original industry. The constituent devices in these systems serve a specific function, but taken together, they become a system that is not merely a collection of individual pieces of equipment. The cybersecurity systems associated with an industrial control system has developed into an ecosystem that would not exist just given the added equipment alone.

Next, I look at a single example of complexity that emerged as a vulnerability through the interactions of functions in the same device. Buffer overflows are an easy programming error to make, easily correctable, and should be caught with proper coding practices. However, the very first instance of a buffer overflow was likely surprising in impact to the developer. Several years later, similar errors in coding made for the possibility of the leaking of private encryption keys in SSL libraries via the "heartbleed" vulnerability. While speculative, I argue that this is exactly what Armstrong and others are alluding to when they write about emergence and complexity in cyber security.

5.2 A Theoretical Example of Emergence Within Cybersecurity

Xu wrote a paper [59] citing the work of [60], where he demonstrates a theoretical argument for the phenomena of emergence produced within a notional cyber system represented as a graph. The argument is based on idea that cyber systems and the interactions between components within the system can be represented as a graph, and the graph can be analyzed with mathematical tools used within dynamical systems, namely eigenvalue and eigenvectors. I find this to be a compelling method to analyze cyber systems for several reasons. Namely, it seems natural to represent cyber systems and their interactions as a graph. In doing so, we are able to map the interactions between components within the system and analyze the how compromising one component could lead to interactions with other components. Within the realm of cybersecurity, this hop-by-hop view of the system would seem to have promise in identifying high value targets and possible paths of compromise. I will recreate Xu's argument here with more detail than he does in his paper. In Chapter 6:, I will extend Xu's work to a generalized electric power SCADA system.

I credit Xu for presenting the idea with a clear and concise explanation, but I will express his argument in a slightly different manner. Consider graphs, G1 and G2, each having an adjacency matrix *$A(G_1)$* and *$A(G_2)$*, respectively. Assume these are two different systems that are independent from each other.

$$A(G_1) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad A(G_2) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Xu references Chakrabarti to explain that the largest eigenvalue of a given graph is related to the systems resilience to epidemic outbreaks. If we look at the eigenvalues of each graph, we see that:

$$\lambda\big(A(G_1)\big) = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \qquad \lambda\big(A(G_2)\big) = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

The eigenvalues are related to the possibility of an epidemic outbreak by:

$$\frac{\beta}{\delta} < \tau \qquad \text{Eq. 5.1}$$

where β is the birth rate of the virus and δ is the death rate of the virus and τ is the epidemic threshold, defined as:

$$\tau = \frac{1}{\lambda_1} \qquad \text{Eq. 5.2}$$

where λ_1 is the largest eigenvalue in the adjacency matrix of the graph.

If the inequality holds—the threshold τ is greater than the ratio of the birthrate to the deathrate—the virus will die out. This means that if the reciprocal of the largest eigenvalue is greater than the ratio of the birth rate to the death rate of the virus, no epidemic emerges.

If we look at Xu's example that we recreated above, we can see that each graph has the same adjacency matrix, and therefore the same largest eigenvalue. For simplicity's sake, we will assume that the virus has the same birthrate and deathrate, and they are independent of the structure of the graph, so β/δ remains the same. Given the eigenvalues of the graphs G1 and G2 above, we can say that if $\beta/\delta < ½$, the virus will die out.

Now, assume that the two systems are connected, and fully interconnected, so that all nodes of G_1 are connected to all nodes of G_2. The adjacency matrix of this graph would take the form:

$$A(G_{1,2}) = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

And the eigenvalues are:

$$\lambda\left(A(G_{1,2})\right) = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 5 \end{bmatrix}$$

We can see that the largest eigenvalue is five. Looking at the systems independently we may have wanted to assume that by adding the two systems, we could add the two largest eigenvalues (which would be four in this example). This turns out to not be the case. Xu uses this to imply that looking at the systems independently and assuming those evaluations will stand when the systems are combined is not true. Something emerges out of the combination of these systems—in this case, $2+2\neq4$, but instead 5, which is a popular definition of the phenomenon of emergence.

5.3 Measures of Complexity

This thesis will not attempt to give an exhaustive list and description of the different measures of complexity—see Seth Lloyd's substantially more exhaustive listing in [46]—but it is important to understand the basic ideas that the great thinkers on this topic have proposed in the effort to quantify complexity. These measures span from the incalculable to the well-defined. We will focus on two examples of computational complexity.

Kolmogorov Complexity

Kolmogorov complexity is an example of a measure of complexity that cannot actually be measured. The idea behind Kolmogorov complexity is to quantify the size of a program used to produce a given computational output relative to the smallest program that could produce that same output [61]. It may be obvious, but the key trouble stopping us from completely quantifying the Kolmogorov complexity of a program is our inability to know what the smallest program possible to compute a necessary output.

Why would such an idea be applicable to cybersecurity? If we have a function, and we would like to code that function into a protective relay, understanding the smallest possible program that could compute that function would allow us to minimize the amount of code that we add to the relay and, in turn, reduce the attack surface or room for coding errors that could lead to vulnerabilities.

I like to think of Kolmogorov complexity as a signal-to-noise problem, or in more accurate terminology, and information-to-noise problem. If we had a program that implemented a function using the smallest possible size program, that program could be thought of to contain pure information. Any program that was larger than the smallest possible program contains code used to contribute to the goal of computing the function but adds unnecessary or extraneous content to the program—in essence, noise.

I can only hypothesize that many reasons for why and how noise enters the picture: ill-suited or poorly constructed programming languages, coders who lack the necessary programming competences, and likely many other conditions.

Computational Complexity

Neil Immerman has an excellent book on computational complexity, titled Descriptive Complexity [62]. He goes on to describe descriptive complexity as a science of looking at programming and computation in terms of both space and time: the space required to host a program and the time required to run the program. From the perspective of space, may be related to the Kolmogorov complexity idea.

Within the idea of descriptive complexity, there are definitions of complexity classes. I will attempt to dive a little into what I perceive as the relevance of complexity classes to cyber security, but I will fully admit the topic of complexity pertaining to complexity classes is a subject far beyond my reach currently. However, I feel there is much to be contributed to cybersecurity from what has already been learned and established in descriptive complexity and complexity classes.

I will restrict my writing on complexity classes to the notion of time required to completely determine a solution and ignore the space aspect of computational complexity. Complexity classes are defined by the time required (or space required, but I am ignoring these classes) to solve a class of programs using specific computing platform, often times a deterministic Turing machine. The most basic classes are defined by programs that run in polynomial time, and nondeterministic polynomial time—P, NP.

5.4 Complexity, Human Error, and Cyber Systems

As we have looked at complexity through the lens of descriptors of hierarchy and emergence, as well as focusing on measures of complexity, it is important to realize that these systems are human made. As such, we must accept that any human designed system is subject to the limitations of human intellect manifested through errors of specification and implementation. In terms of the science of human error, these concepts can be thought of as mistakes and slips [63].

However, before we dismiss the idea of cybersecurity as a hopeless cause because of human fallibility, we should spend some effort understanding a few of the mechanisms that scientists and engineers have developed to limit the probability of errors that could impact the security of a cyber system.

Perhaps Bell and Lapadula were on to something when they started their investigation of formalizing confidentiality using the general system theory (GST). GST is the precursor theoretical framework to complexity science, which, starting in the 1980s, became a more widely accepted framework for understanding the interrelations of systems [47]. Heylighen et al. describe the idea of complex systems in an easily understandable way:

> As technological and economic advances make production, transport and communication ever more efficient, we interact with ever more people, organizations, systems and objects. And as this network of interactions grows and spreads around the globe, the different economic, social, technological and ecological systems that we are part of become ever more interdependent. The result is an ever more complex "system of systems" where a change in any component may affect virtually any other component, and that in a mostly unpredictable manner. [47]

The thrust of their argument is that complex systems are not well-suited problems for classical Newtonian physics or analysis. Newtonian mechanics depend on a closed system. Problems solved in using these tools are confined to relatively simple interactions of finite sets of objects that are well defined in the sense that we understand their relationship within the system. Even then, when modeled too granularly, Newtonian mechanics break down.

A good example of this breakdown is the n-body problem in calculating the gravitational attraction between n-bodies in a system. Newton's law of gravity works well when we just consider two bodies interacting with each other. However, after adding a third body or more, the equations break down. In fact, there doesn't appear to be a general analytical solution to the n-body problem [64]. This is likely because of the seemingly infinite number of objects that are acting on the bodies in question. To accommodate for all interaction would at least be algorithmically unfeasible.

We appear to run into a similar problem in cybersecurity. The sheer number of devices that interact with each other in a cyber system is likely unknowable. The sheer complexity of the system stretches our ability to analyze it using traditional tools. This raises the question: are there other sets of tools that are more applicable?

With this question, we turn to complexity science. Above, I demonstrated, using Xu's proof, that emergence within cyber systems is possible. With this, we will look at how we can use complexity science to formally proof an aspect of cybersecurity. In Chapter 6 we will develop a method to prove relative security amongst various cyber systems using power system protection and control systems.

Chapter 6: Notional Cyber Security First Principles Applied to Power System Protection

6.1 Protection Systems, Uncertainty, Trust, and Complexity

As I try to pull all of the concepts discussed in the previous chapters together to create some semblance of a coherent formalism, I feel the most promising topic in the discussion that can extend to a first principle is complexity. As detailed in earlier chapters, uncertainty is the most fundamental of all of the concepts that pertain to cybersecurity. I outlined the argument that reducing uncertainty increases information. However, in our discussion of complexity, I highlighted the idea that a system of perfect information does not exist; hence, we will always be subject to a signal to noise problem. With this, I looked at the role that complexity plays in the signal to noise ratio, and I argued that reducing complexity is tantamount to reducing noise in the system. As such, complexity jumped out in my research as the most promising concept able to be formalized and used in analysis of the cybersecurity of protection systems.

While the rest of this chapter will discuss how reducing complexity can be used to prove relative security of cyber systems, I would like to advocate that the time spent discussing the concept of trust was certainly not a waste of effort. It should be obvious that the use of trust involves a great amount of uncertainty and, subsequently, complexity. Given this, I expect that the results that I present throughout the rest of this chapter could be applied to the notion of trust as well. I will leave this discussion to Chapter 7 and discussion of future work.

6.2 Formal Proof for Optimizing the Security of Protection System Architectures Through Reductions in Complexity

As we look at the complexity of protection systems, the thought of reducing complexity to improve security seems obvious. However, I was not able to find much literature that defined formal methods to support complexity reduction to improve system security. As outlined above, Shouhuai demonstrated the theoretical possibility of emergence in complex system in [59], which is related to the notion that reducing complexity can improve security. My interpretation of Shouhuai's argument was that even if we had perfect knowledge of two systems, A and B, if we connect them, properties that we may not have been able to predict analyzing the systems in isolation will emerge. This seems to suggest that if we can reduce the complexity of systems, then we will lessen the chances or frequency of emergent behavior. I can only imagine this would be a net positive for the cybersecurity of a protection system.

Interestingly, in [59] Shouhuai cited [60], which argues that complex systems, like cyber systems, can be modeled using epidemiological models. In the text below, I will argue that we can use these models to formally prove that a reduction in complexity can improve system security. To lay the groundwork for this claim, I will review recent research in epidemic processes and discuss how they relate to complex systems like cyber systems.

Chakrabarti et al. describe in [60], and reviewed in [65] by Pastor-Satorras et al., that the conditions in which a virus will break out in a population if the ratio of the birth rate β of a virus to the deathrate δ exceeds a threshold τ.

$$\frac{\beta}{\delta} > \tau \qquad Eq\ 6.1$$

Chakrabarti goes on to prove that the threshold τ is equal to the reciprocal of the largest eigenvalue of the adjacency matrix of the graph that represents the network on which the virus is propagating.

$$\tau = \frac{1}{\lambda_1} \qquad Eq. 6.2$$

I detail Chakrabarti's proof in Appendix A and his original argument can be found in [62]. Suffice it to say, regardless of the birth or death rate of the virus, Chakrabarti's finding implies that the lower the largest eigenvalue of a network graph adjacency matrix, the higher the threshold to an epidemic condition occurring in the network. In other words, any reduction in the largest eigenvalue of the network graph adjacency matrix increases the resistance of the network to an epidemic outbreak.

Considering this finding, deriving a mathematically provable relationship between a network graph and its largest eigenvalue would allow us to provably compare the relative security between network architectures. This notion led me to ask, "So what is that relationship?"

As is turns out, there exists a relationship between the largest eigenvalue of a graph and the number of edges between vertices on the graph. The fewer the edges, the lower the largest eigenvalue. Using a specific example to illustrate, consider the two simple graphs below, G_1 and G_2:

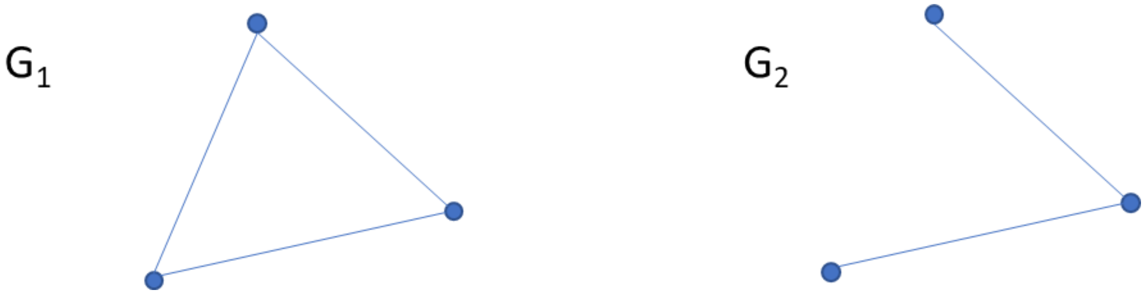

Figure 6.1 Graph G1 with three edges and G2 with two edges

If we create adjacency matrices for the graphs (*A*) and perform eigen-analysis we can compare the differing largest eigenvalues (λ).

$$A_{G1} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$\lambda_{G1} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$A_{G2} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\lambda_{G2} = \begin{bmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}$$

Just using this simple example, we are able to see the conjecture above holds. Graph G2 has fewer edges, and the largest eigenvalue is smaller than G_1 (2 for G_1 and √2 for G_2). Now we need to ask, "Is this result generalizable?" Does it hold for all graphs characteristic of networks in cyber systems?

I provide the proof below to support the conjecture above. The proof appears in [66]; however, I added supplemental information to hopefully support the proof and make it more understandable. In [66], Cvetkovic et al. provide the following proposition:

"If G-uv is the graph obtained from a connected graph G by deleting edge uv, then λ_1(G-uv) < λ_1(G)." [66]

We will take "edge uv" to be the edge connecting vertex u and vertex v, λ_1*(G-uv)* to be the largest eigenvalue in the adjacency matrix of graph *G-uv*, and λ_1*(G)* to be the largest eigenvalue in the adjacency matrix of graph *G*.

Cvetkovic, et. al.'s proof follows:

> Let $\boldsymbol{x}$= (x_1, x_2, …, x_n)⊤ be a nonnegative unit eigenvector of *G-uv* corresponding to *λ_1(G-uv)*. Then
>
> $$\lambda_1(G-uv)=\boldsymbol{x}^\top A(G-uv)\boldsymbol{x}\le\boldsymbol{x}^\top A(G)\boldsymbol{x}\le\lambda_1(G)\qquad Eq.6.3$$
>
> If *λ_1(G-uv)* = *λ_1(G)*, then $\mathbf{x}$ is the principal eigenvector of *G* and hence has no zero entries. Now
>
> $$\boldsymbol{x}^\top A(G-uv)\boldsymbol{x}=\boldsymbol{x}^\top A(G)\boldsymbol{x}-2x_ux_v<\lambda_1(G-uv)\qquad Eq.6.4$$

This inequality is a contradiction.

This proof demonstrates a contradiction that is created if we assume that two graphs, both identical except for G2 has one edge removed, can maintain at least equal values for their largest eigenvector. I will step through the proof piece by piece to make sure that everyone, especially me, can understand it.

$\boldsymbol{x}$ is the unit eigenvector of the largest eigenvalue (the notation λ_1 generally denotes the largest eigenvalue of a given graph) of graph *G-uv*. It is important that we use the unit eigenvector in this proof, otherwise the numbers will not work. The term "non-negative" is important as well. The non-negative nature of the eigenvector is a characteristic of the largest eigenvalue of undirected graphs. This is not the case for all matrices but specific to the adjacency matrix of undirected graphs.

The first statement with inequalities states that the largest eigenvalue of *λ_1(G-uv)* is equal to the transpose of the eigenvector associated with the largest eigenvalue of *G-uv*, multiplied by the adjacency matrix of *G-uv*, multiplied by the eigenvector associated with the largest eigenvalue of *G-uv*. The first eigenvector is transposed so that it can operate on the adjacency matrix. The matrix operations produce a single value. This single value, according to the inequality, is less than the largest eigenvalue of *G-uv*, multiplied by the adjacency matrix of *G*, multiplied by the eigenvector associated with the largest eigenvalue of *G-uv*.

Note that the largest eigenvalue of *G-uv* is operating on the adjacency matrix of *G*, not *G-uv*. These matrix operations produce a single value that is less than the largest eigenvalue of *G*.

The last two lines of the proof set the stage for the contradiction that arises. We assume that we can set the largest eigenvalue of *G-uv* equal to the largest eigenvalue of *G (*λ_1*(G-uv)*= λ_1*(G))*. This means that we removed an edge from *G* to create *G-uv*, but the two graphs still have the same largest eigenvector. Also note that because the eigenvector $\boldsymbol{x}$ is the eigenvector of the largest eigenvalue, it has no zero entries. This becomes an important characteristic which is required for generalization of all similar graphs.

The largest eigenvalue of *G-uv (*λ_1*(G-uv))* has the following equality:

$$\lambda_1(G-uv) = \boldsymbol{x}^\top A(G-uv)\boldsymbol{x} \qquad Eq. 6.5$$

Since G-uv is equal to G minus one edge, we can rewrite the equality in terms of G

$$\lambda_1(G-uv) = \boldsymbol{x}^\top A(G)\boldsymbol{x} - 2x_u x_v \qquad Eq. 6.6$$

But, if λ_1(G-uv) = λ_1(G) and $\boldsymbol{x}^T A(G)\boldsymbol{x}$ was at best equal to λ_1(G), stating that $\lambda_1(G\text{-}uv)= \boldsymbol{x}^T A(G\text{-}uv)\boldsymbol{x}= \boldsymbol{x}^T A(G)\boldsymbol{x}\text{-}2x_u x_v$ is a contradiction. If $\boldsymbol{x}^T A(G)\boldsymbol{x}$ was at best equal to $\lambda_1(G)$, how can $\boldsymbol{x}^T A(G)\boldsymbol{x}\text{-}2x_u x_v$ still be at best equal to $\lambda_1(G)$? $\boldsymbol{x}^T A(G)\boldsymbol{x}$ minus positive number must be less than $\boldsymbol{x}^T A(G)\boldsymbol{x}$, hence the contradiction.

An important part of this proof lies in the statement $\boldsymbol{x}^T A(G\text{-}uv)\boldsymbol{x} = \boldsymbol{x}^T A(G)\boldsymbol{x}\text{-}2x_u x_v$. Where does the -2 $x_u x_v$ come from?

To answer this, we will examine the matrix operation $\boldsymbol{x}^T A(G)\boldsymbol{x}$.

For graphs G1 and G2 in Figure 6.1, the associated matrices are:

$$A_{G2} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_{G1} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$\lambda_{G2} = \begin{bmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix} \rightarrow \lambda_{1,G2} = \sqrt{2}, \qquad \lambda_{G1} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \rightarrow \lambda_{1,G1} = 2$$

$$x_{1,G2} = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{bmatrix}, \qquad x_{1,G1} = \begin{bmatrix} 0.577 \\ 0.577 \\ 0.577 \end{bmatrix}$$

$$\boldsymbol{x}_{1,G2}^\top A(G)\boldsymbol{x}_{1,G2} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{bmatrix}$$

To understand the -2$x_u x_v$, consider that x_u and x_v are the elements of the $\boldsymbol{x}$ vector that are multiplied by the elements in the adjacency matrix that went from 1 to 0 when the edge was removed. In graph notation, they are elements $x_{1,3}$ and $x_{3,3}$ of **x**. Dropping the subscripts in the notation going forward, graphically,

$$\boldsymbol{x}^{\top}A(G)\boldsymbol{x} = \begin{bmatrix} \underbrace{\frac{1}{2}}_{x_u} & \frac{\sqrt{2}}{2} & \underbrace{\frac{1}{2}}_{x_v} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \overbrace{\frac{1}{2}}^{x_u} \\ \frac{\sqrt{2}}{2} \\ \underbrace{\frac{1}{2}}_{x_v} \end{bmatrix}$$

Figure 6.2 Identification of x_u and x_v

The $-2x_ux_v$ terms is included in the equation above because in order to equate $\boldsymbol{x}^TA(G\text{-}uv)\boldsymbol{x}$ with $\boldsymbol{x}^TA(G)\boldsymbol{x}$, we need to account for the removed edge between vertices u and v. These vertices are highlighted in the *A(G)* matrix in Figure 6.2.. In *A(G-uv)* these highlighted elements are zero. Therefore to get from $\boldsymbol{x}^TA(G)\boldsymbol{x}$ with $\boldsymbol{x}^TA(G\text{-}uv)\boldsymbol{x}$, we need to subtract the impact these elements have. Carrying out the calculations, we get:

$$\boldsymbol{x}^{\top}A(G) = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} + \frac{1}{2} & 1 & \frac{\sqrt{2}}{2} + \frac{1}{2} \end{bmatrix}$$

And therefore

$$\boldsymbol{x}^{\top}A(G)\boldsymbol{x} = \begin{bmatrix} \frac{\sqrt{2}}{2} + \frac{1}{2} & 1 & \frac{\sqrt{2}}{2} + \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{bmatrix}$$

Solving the full $\boldsymbol{x}^TA(G)\boldsymbol{x}$ we see that x_u and x_v are multiplied by each other twice as a result of the edge remaining in the adjacency matrix, which is why we need to subtract $2x_ux_v$ from $\boldsymbol{x}^TA(G)\boldsymbol{x}$ to arrive at $\boldsymbol{x}^TA(G\text{-}uv)\boldsymbol{x}$.

To summarize, using Cvetkovic's proof, I am able to justify the claim that certain network structures are more resistant to propagation of viruses than others. This statement allows us to look at network architectures to determine relative security of comparable network topologies. This is accomplished by examining the largest eigenvalues of comparable networks, where the architecture whose first eigenvalue is the smallest is theoretically more resistance to virus propagation.

As I will show in the next section, this finding allows us to architect networks in measurably more or less secure ways by understanding the network graph and working to reduce the largest eigenvalue. We are able to reduce the magnitude of the first eigenvalues, for instance, largest, of graphs by physically or logically reducing the number of edges connecting vertices within the adjacency matrix of the graph.

6.3 What Does this Formalism Mean for the Design of Protection System Networks?

Consider the image in Figure 5.3 and the corresponding graph, Figure 6.3 below. We can build the adjacency matrix for the graph in Figure 6.3. For simplicity's sake, we will only focus on the section in Figure 6.3 circled in green. We could do the entire graph in a similar manner if desired.

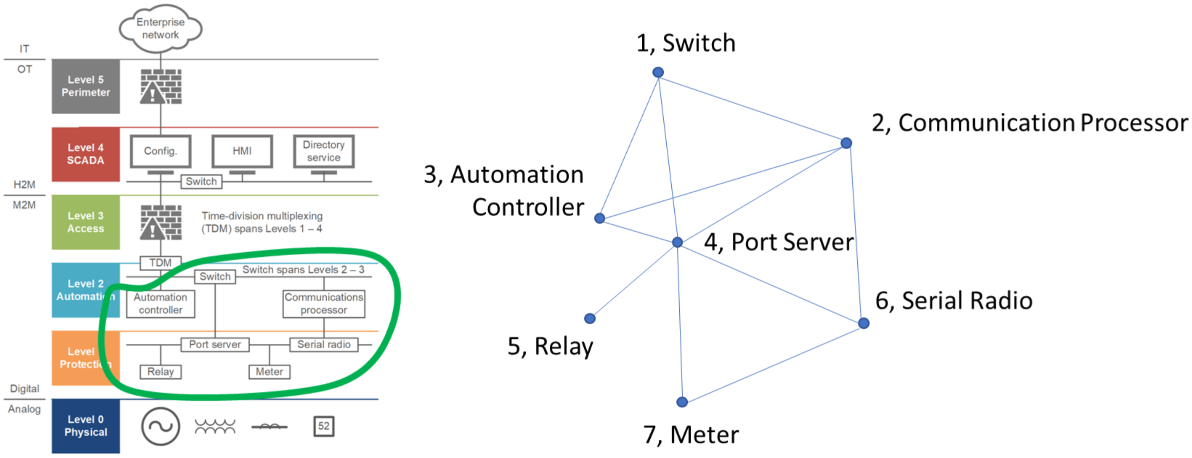

Figure 6.3 Network topology and associated graph

A(G_N) represents the adjacency matrix of graph G_N.

$$A(G_N) = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Solving the adjacency matrix for the eigenvalues, we get:

$$\lambda\big(A(G_N)\big) = \begin{bmatrix} -1.9 \\ -1.7 \\ -1 \\ -0.5 \\ 0.3 \\ 1.1 \\ 3.6 \end{bmatrix}$$

We see that the largest eigenvalue of the adjacency matrix $\lambda_1\big(A(G_N)\big) = 3.6$. If we re-architect the network, and reduce the number of edges, we can reduce the largest eigenvalue of the graph. Assume we remove the physical connection between the switch and the port server and communication processor. We will assume that the automation controller must be compromised to allow propagation of malicious activity downstream to the port server and automation controller. The resulting graph takes the form:

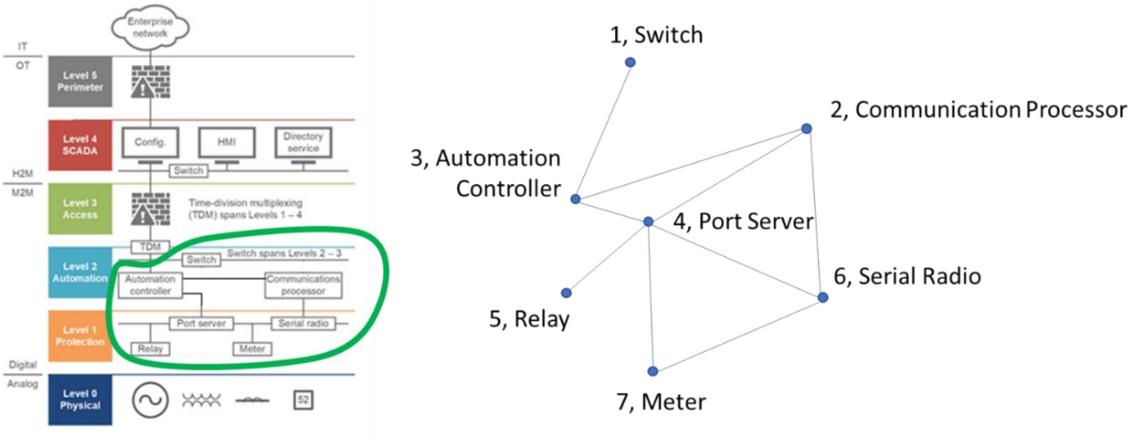

Figure 6.4 Reduced-edge network architecture

The associated adjacency matrix and its eigenvalues are:

$$A\big(G_{N\_New}\big) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$\lambda\big(A(G_N\_New)\big) = \begin{bmatrix} -1.9 \\ -1.7 \\ -0.9 \\ 0 \\ 0.3 \\ 1.0 \\ 3.1 \end{bmatrix}$$

We can see that the largest eigenvalue of $\lambda\big(A(G_N\_New)\big)$ is smaller than that of $\lambda\big(A(G_N)\big)$. Per Chakrabarti's findings regarding the propagation of viruses, $\lambda\big(A(G_N\_New)\big)$ presents a theoretically stronger architecture to resist malicious attacks.

It is important to understand the differences between various logical separations. In terms of propagation of virus in the medical sense, the elements of the adjacency matrix of the graphs—which depict transmission paths of viral propagation—that have a value of one represent direct, physical interaction between the vertices. In the case of protection systems, or any other system comprised of electronic devices, we need to be careful of we think of these interactions. Physical separation does not necessarily imply logical separation. Logical separation is very important for our contexts. The graphs of cyber systems must include both physical and logical connections.

Physical separation is easy to understand. Devices that are physically separate do not have a direct communications path between themselves. Those connections might be fiber optic, wireless, or any other electromagnetic transmission media. However, logical connections are a bit trickier. If a network switch passes information from one device to another, completely unencumbered, it may not be a direct physical connection—in the sense of a single, continuous copper wire trace from input to output—but surely there is a logical connection that presents a similar transmission path for a virus, with little encumbrance.

Thinking about these logical connections, we can break down the various types of logical separation in terms of the different types of network appliances that are connecting vertex A to B. From least to most configurable we have:

1. Hub
2. Unmanaged switch
3. Managed switch
4. Layer-3 switch
5. Software-defined network (SDN) switch

At one extreme, we have a hub: a pure physical-layer routing of incoming traffic on one port to all other ports. Anything received on one port gets forwarded to all other port. This would be the same as a direct physical connection between two devices. On the other extreme, we have a software-defined network switch. A software-defined network switch must be programmed to allow every type of packet that the downstream device will be expected to see. This represents a logically separable device that can be thought to sever the connection between two devices, much like removing a physical connection would sever two devices.

The intermediary devices have some room for interpretation. Can the unmanaged switch be thought to sever that physical connection between upstream and downstream devices? The answer to that question could be debated. However, the further down the list we go, the better argument we will have to treat the device as a physical separation between upstream and downstream. This is especially true if we start thinking about the links at a lower level.

Consider a layer-3 switch and the SDN. The layer-3 switch can be configured to filter IP address passed to a particular downstream connection, a form of logical separation. However, any valid IP address would be able to pass all traffic downstream. The SDN switch however, could be configured to only allow a specific protocol to pass from the upstream to downstream, another form of logical separation, but in a stricter sense. To account for these cases of logical separation, we may want to consider expanding the graph to include not only physical separations but the logical separations as well.

While this finding supports Chakrabarti's claims, the findings should not be considered conclusive. Van Mieghem et al. suggest such threshold conditions are true in steady-state [67], and present a Markov theory-based model to shows the same system developing an absorbing state. With this, Van Mieghem, et. al. claim to show that Chakrabarti's model is only accurate when the virus spreading rate is below the threshold. Despite this result, it does seem reasonable to hold to the general idea of this thesis in that reducing the spectral radius of the graph will still have a net positive effect on the resilience of the architecture to an attack. Van Mieghem also goes on to support the notion of reduced spectral radius to improve the threshold to epidemic outbreaks in later papers [65] [68].

In [69], Xu et al. cites Chakrabarti's findings to point out that there is room for improvement in the model that Chakrabarti uses. Xu recommends a more detailed model that includes arbitrary dependencies when an attack is taking place. Neither Chakrabarti nor Van Mieghem include any sort of dependencies within the graph, though, Xu argues, cyberattacks likely have certain dependencies while they are being executed. However, Xu's work does still seem to support the notion that there is a relationship between the spectral radius of a graph (i.e., the largest eigenvalue of the graph) and the resilience of that topology to a cyber-attack.

All of this is to say that there are likely improvements that can be made to the models above and those from literature that are not explored here. We seem to be on the precipice of a much more sophisticated understanding of the dynamics of a cyberattack using epidemiological models.

6.4 The Tension Between Power System Protection and Cybersecurity

Ultimately, protection is not implemented at the service of cybersecurity; it is the other way around. The protection of the system is the primary concern and cybersecurity needs to allow the protection to operate as required. However, without proper implementation of adequate cybersecurity practices, a microprocessor-based protective relay may be left vulnerable to malicious actors—the worst case being when a relay has been clandestinely compromised and will not operate as expected under given conditions.

In nearly all cases, protective relays do not require communication between devices to operate. The obvious exception is the differential relay that requires relays at remote ends of power lines to exchange time-aligned current measurements to properly calculate restraint and operate quantities. Aside from this, protective relays do not require communication between devices to protect a system, but modern implementations of protection schemes have used communications to improve selectivity, speed, and availability of systems. Availability has likely been the single largest attribute of a system where improvement depends on the expanded deployment of sophisticated communications systems, and these systems can be considered cyber systems.

Herein lies the problem: protection systems do not, strictly speaking, require communications to operate satisfactorily, but advancements in communications have increased the capability of protection systems to the point where some systems are designed around communications capabilities. To this end, the most secure form of protection, differential protection, does require communications between relays. All of these communication technologies require their own set of cybersecurity protections, but traditional cybersecurity tools can hinder the performance of the communication system, consequently degrading the performance of the protection system. The introduction of the cybersecurity risks as a consequence of these communications now puts protection and automation system designers in the unenviable position of balancing the need to ensure that the protection system operates as required for the system while also securing the protection- and control-specified communications systems from cyberattack. As I will explain, these two requirements can sometimes be in tension.

Protection systems require determinism, speed, and high availability. Traditional communications security technology tends to not focus on determinism and focuses on speed in terms of bandwidth but not necessarily speed in terms of latency—for example packets per second may not be the focus, rather bits per second, which is maximized by using fewer, but larger, packets. Availability is traditionally a focus of communication systems; however, the lack of focus on speed (reduced latency) as a requirement makes the tools used for increased availability less useful than desired for a protection system.

For example, consider a zone interlocking protection scheme. At a high-level, the scheme permits fast coordination between protection devices using communication channels for devices to identify where a fault is and have the appropriate device protect the system against that fault, while sending blocking signals to the devices that also see the fault but need not trip. Traditional methods of protection would use predetermined time delays for each device, but the communication channel reduces the delays required by allowing the devices to communicate directly to each other. This speeds up the protection and minimizes the damage done to the assets being protected.

Consider Figure 6.5 which was created based on ideas and illustrations presented in [70]. This presents a simple zone interlocked main-tie-main power system bus configuration (bottom diagram in the figure) and corresponding network diagram (top diagram in the figure). Switches 1-3 are linked together in a ring architecture. The switches use a technology call rapid spanning tree protocol (RSTP) to decide how the traffic flows between the switches. If a link between switches fails, the switches must decide how the traffic must be rerouted and this takes time. RSTP is designed to improve the availability of a system, however the time for the switches to converge to a new traffic flow pattern can be insufficient for critical systems. Where protection related critical networks often demand sub-second convergence times, RSTP networks can take several seconds to upwards of a minute or more depending on the size and architecture of the network.

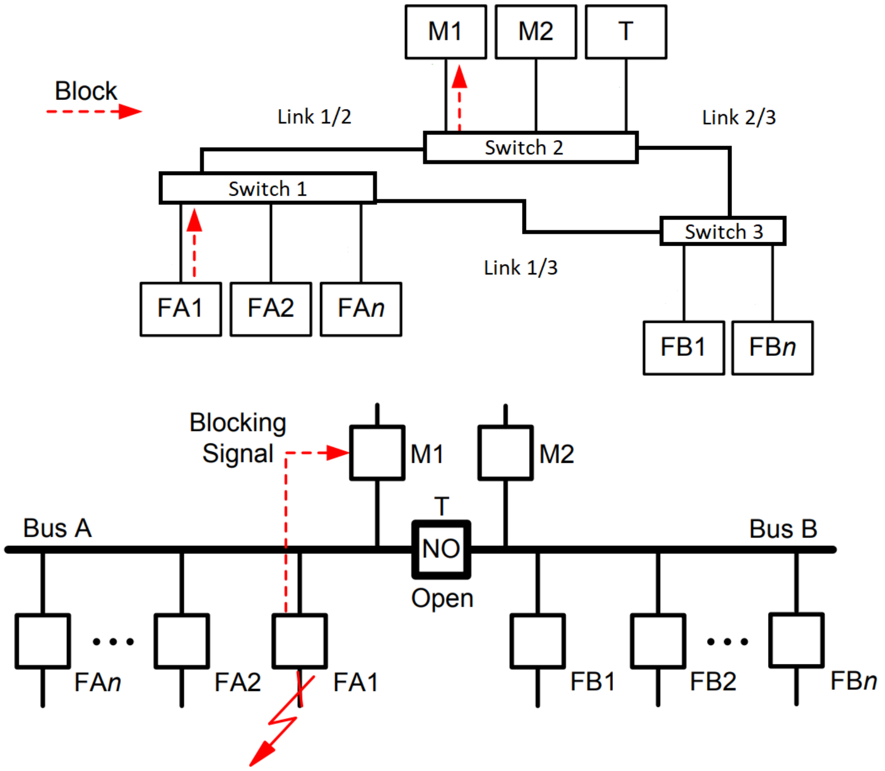

Figure 6.5 Zone interlocking system example

In this example, a fault occurring below breaker FA1 initiates the protective relay controlling breaker FA1 to send a block signal to breaker M1. Opening breaker FA1 isolates the fault completely, whereas opening breaker M1 also isolates the fault, but impacts the loads downstream of FA2-FA*n*. Such a system allows the protective relay controlling M1 to operate with a only a small time delay because the downstream protective relays act quickly to send blocking signals to the M1 device if they (downstream relays) detect a system fault in their zone. In this example we increase the speed and selectivity of the protection system but add a communication system that may open the door for cyber-related attacks.

The notional system represented in Figure 6.5 serves as a good example to illustrate the tension between protection and cybersecurity. According to the observation that reducing the number of edges in a network increases the threshold of a network to a cyberattack, the system represented in Figure 6.5 would be best served by removing one of the links, as shown in Chapter 5 and in Figure 6.6. However, if we remove one of the links, we suffer from introducing a single point of failure in the communications system, thereby reducing its availability. And, removing the communication system altogether decreases the speed and selectivity of the protection systems, as I mentioned above.

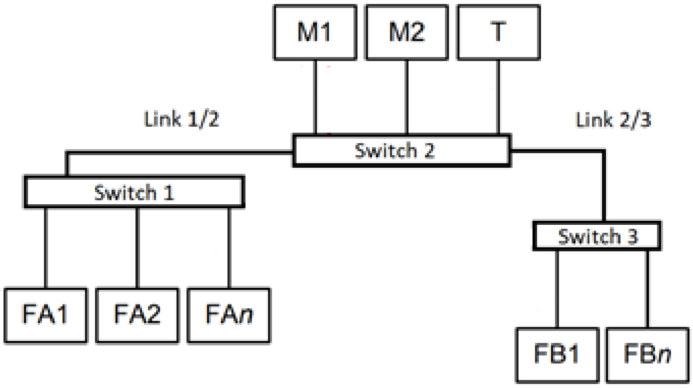

Figure 6.6 Reduced-edge communication network zone interlocking example

In such a case, improving availability to the protection systems tends to stress the protection/cyber system relationship. As mentioned above, improving system availability often requires adding communication systems. A very specific example is the addition of redundancy to the protection systems. This is very prevalent in systems where voting schemes are employed. The act of adding redundancy to the system adds additional devices and/or communication networks to the system to maintain operational status in situations where one system or device may fail.

I can point to another example in more recent protection and SCADA system networking technology that improves network availability using parallel redundancy protocol (PRP) or high-speed redundancy protocol (HSRP). These protocols were developed to improve network re-convergence times under device or link failure scenarios, recognizing RSTP deficiencies. The idea is to improve availability, but this requires adding more devices and more links between devices. See Figure 6.7, which illustrates a notional network that represents a generic system using a PRP- or HSRP-like technology. The switches (e.g., Sw1A, Sw2B, etc.) are connected in two independent rings, the A ring and B ring. Each device (D1-D3) has two network connections and are connected to both ring A and ring B.

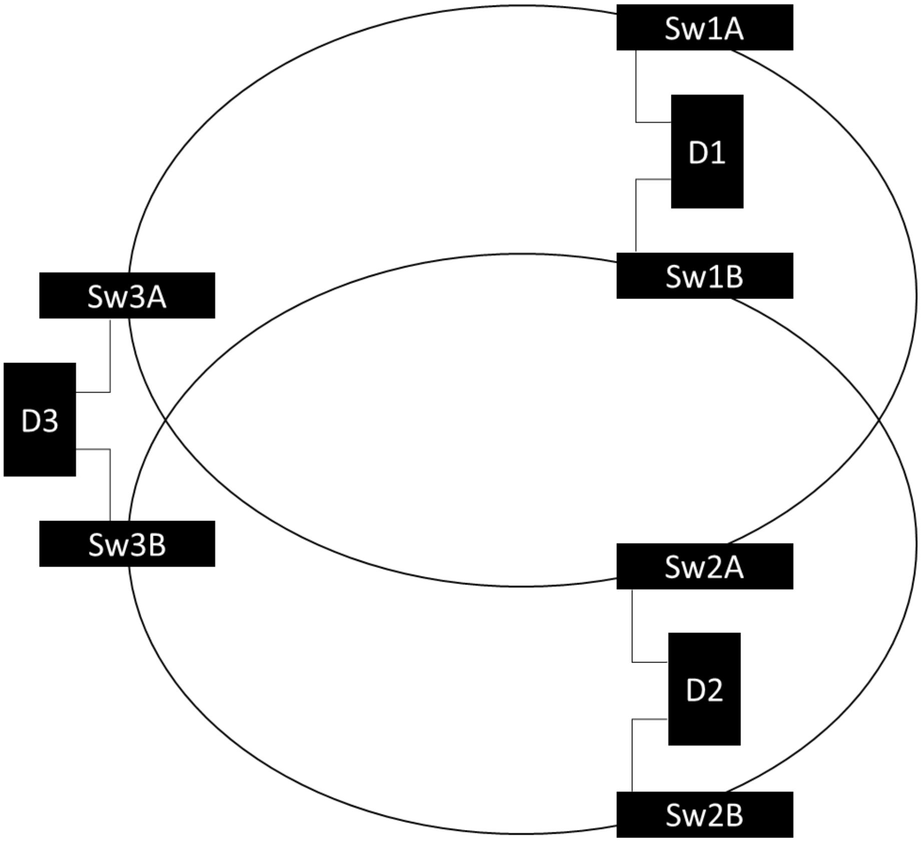

Figure 6.7 Notional PRP- or HSRP-like network architecture

Looking at this system, we can recognize a similar graph representation of the system. Since the communication rings are separate except at the connection with each device, the networks are not fully connected (i.e., every vertex connected to every other vertex). The connection graph, associated adjacency matrix, and resulting eigenvalues of the graph in Figure 6.8 are shown below.

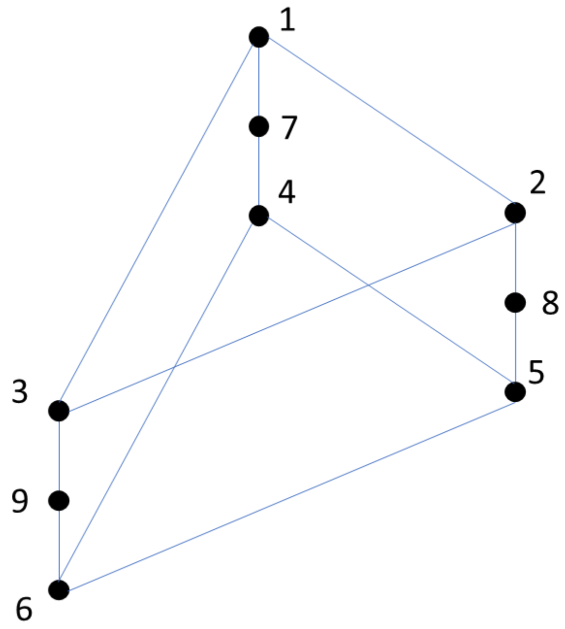

Figure 6.8 Graph representation of the notional PRP or HSR network

$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\lambda\big(A(G)\big) = \begin{bmatrix} -2 \\ -1.6 \\ -1 \\ -1 \\ 0 \\ 0 \\ 1 \\ 2 \\ 2.6 \end{bmatrix}$$

We can begin to look at this increase—one network ring versus two—and evaluate if the marginal increase in the largest eigenvalue is worth lowering the threshold of resilience to a viral outbreak, as demonstrated in Chapter 5. Figure 6.9 shows a single ring and the associated graph, adjacency matrix and eigenvalues. The difference in the spectral radii of the two graphs is 0.2, or less than 8%. While this thesis has not presented ideas to judge absolute measures of security based on graph spectral radius, a relative difference of 8% does not represent a major difference. So, to address the question starting this paragraph in a slightly different way, is the marginal increase in threshold to viral outbreak worth the reduced network failover time?

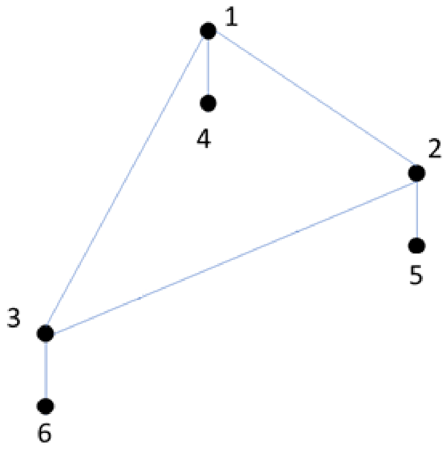

Figure 6.9 Single ring system

$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\lambda\big(A(G)\big) = \begin{bmatrix} -1.6 \\ -1.6 \\ -0.4 \\ -1 \\ 0.6 \\ 0.6 \\ 2.4 \end{bmatrix}$$

Such an evaluation should consider whether or not the reduced resiliency of the system comes with an increase in protection system availability or simply increased availability for the SCADA system. If we consider the protection system to be the primary system and design the communication system around the needs of the protection system, then adding resilience to the protection system may be worth added cyber risk. However, adding more cyber risk for enhanced SCADA capability does not serve the primary objective of protecting the power system and should be heavily scrutinized.

Another point to consider is that adding redundancy irrefutably increases the attack surface of the system, but how important is this, or maybe more appropriately, when is this important? If the two ring networks are identical, is there any disadvantage to having two identical networks? Even conservatively assuming that the network is exploitable because a device on the network, a switch for example, has a vulnerability, I am not sure that having two devices that are identically exploitable makes much difference in the overall security of the system. That said, I would submit that it would likely be a disadvantage if the networks were using different vendors/manufacturers for ring A versus ring B in Figure 6.7. This would give an attacker two different manufacturers to target, approximately doubling their chances of finding an exploitable vulnerability. None of this is addressed through the eigenvalue analysis of the network connections graphs. So, while the graphs and their subsequent analysis can provide us with relative system metrics for system security, it is by no means comprehensive in its coverage of all system cybersecurity concerns.

Going back to redundancy and its impact on cybersecurity, it is common for protection systems to have primary and backup protection. Many high-voltage transmission lines in the United States have two protective relays at each end of a transmission line. This is an example of redundancy in that if one relay fails while in service, protection availability on the line remains healthy because a second protective relay is installed and active. A malicious actor now has two devices in which to try to compromise to gain access to the system. Also, this redundancy seemingly goes against the argument presented above that reducing the number of edges of the system improve the cybersecurity of the system. How do we reconcile this conflict?

Considering all of this, the question becomes: how do we balance the design of protection systems and cyber systems so that the protection system does its job in minimizing damage to equipment when faults occur, but also design a cyber system that protects the protection system and remains resilient to cyber intrusion and malicious actor threats? The answer is unclear. However, I remain convinced that the cyber system is at the service of the protection system. As pointed out earlier, bolstering the protection system may increase the attack surface of the system, but at least this risk stems from improving the primary purpose: maximizing the ability of the system to protect the assets under its purview. With this line of thought, I argue that the protection system should be designed as needed to satisfy the level of protection required, and from there, the cyber system should be minimized to accommodate the protection system, and not reduce it to a subordinate system.

Chapter 7: Conclusions and Recommendations for Future Research

7.1 Summary of Findings

I have shown that we can formally prove the relative security of one system design over another using eigenvalue analysis of the associated graphs of the systems. While this analysis cannot provide us with an absolute measure that we would expect from a traditional first principles-type calculations, it advances us in the right direction.

7.2 Conclusions

There are some obvious unanswered and unsatisfactorily answered questions when using this eigenvalue analysis. The most significant in my mind is that while the analysis applies to human social network interactions and information technology systems, how appropriate is it for operational technology related systems that do not share all of the same characteristics of the either? There are some unanswered questions about how redundancy plays into the performance characteristics of the system and how to account for these characteristics using the eigenvalue analysis. And this analysis focuses on only one aspect of cybersecurity of the connections between devices. Such a narrow focus is certainly insufficient to address the cybersecurity of the whole of systems. That said, there are no shortage of opportunities for those that are interested in exploring the first principles of cybersecurity.

The JASON program group at MITRE Corporation states in a report [71] that "There are no intrinsic "laws of nature" for cyber-security…" I find this to be an incredibly bold statement considering the relatively short time that we have been trying to address the challenges of cybersecurity. Throughout this thesis, I have argued that there are several fundamental principle-like concepts that seem to be related to the search for first principles of cybersecurity: the fundamentals of uncertainty as exemplified by Shannon's theorem for bandwidth required to transmit information in the presence of noise, the notion of complexity, and the reduction of complexity to improve the security of a cyber system. Complexity extends to the idea of complexity classes and the theory behind complexity classes and the $P \neq NP$ research, and likely others.

7.2 Recommendations for Future Research

On a practical level, continued investigation into the modeling of cyber-attacks using epidemiological models seems to hold great promise. This work extended the work of others to prove the ability to increase the threshold of a system to an attack by altering the configuration of the network and working to reduce the number of logical connections between devices. Can we create even more rigorous controls? How do the epidemiological models differ from the models of cyber-attacks? Can we create better models that represent more sophisticated attacks that have a fixed motivation to disrupt a process, versus the somewhat arbitrary spreading of a virus through a system? Does this same reduction of edges in the system graph continue to increase the threshold to resisting these types of attacks? What is the impact to the eigenvalue method of adding redundancy, and how do we quantify redundancy for protection purposes versus redundancy for automation or remote monitoring purposes? Also related, what impact does multiple vendor solutions have? Is it better cybersecurity-wise to have a single vendor solution, and can the eigenvalue analysis be used to prove or disprove this?

More theoretically, the concepts presented above may very well represent laws of nature where we cannot answer the question "Why?" but may begin to harness an understanding of "How?" Jeff Erickson mentions the P vs. NP problem in his book "Algorithms" and speaks to it potentially being a law, just like Maxwell's equations [72]. In Maxwell's equations, we don't understand why a rotating magnetic field produces a voltage, or the divergence of a magnetic field is zero, but that is exactly what we observe at the macro-scale. Perhaps $P \neq NP$ falls into the same category. We don't understand why it is a law, but nothing that we have observed thus far contradicts its premise.

For me, the P and NP question, and further exploration into the field of complexity in general, holds some sort of promise to help us understand first principles of cybersecurity. While in this thesis I proved that a reduction of the complexity of system architecture can be shown to improve the cybersecurity of a system, there are many other angles to explore within the concept of complexity. Many questions stem from the basic premise of the P versus NP complexity argument: there exists a set of decision problems where the solution is easily verified if you already know the answer, otherwise verifying that a solution exists in intractable.

It would be interesting to understand if cybersecurity has some fundamental relationship to the P versus NP from the standpoint of verifying that the system of systems that is cybersecurity contains no vulnerabilities. Like decision problems in the NP space: is verification only possible if you already know the answer, thereby making any effort to computationally evaluate a system for vulnerabilities impractical? Or is the problem more similar to a decision problem in P where there is a computationally efficient algorithm to verify the security of a system?

Another prime area for investigation is the role of trust in cybersecurity. As technology expands, it seems that the general population is extending trust to technology far beyond reasonable limits. Consequently, that trust is, from my point of view, being violated at alarming rates. With trust being essential to hold the very fabric of society together, what does it mean for the future of society if these violations of trust continue? I argue that cybersecurity and improving cybersecurity plays and existential role in society, not just to keep the lights on and provide clean water, but to the very core of safeguarding humanity.

There is no shortage of problems to solve in the domain of cybersecurity. I am happy being involved in the conversations and submit this thesis as a way to promote conversations, prompt debate, and encourage others to think beyond current technology. Developing new widgets to solve the most challenging problems in cybersecurity using the same technology on the same platforms is, at least in part, responsible for putting us in the situation that we now find ourselves. I believe we will only succeed in creating secure cyber systems when we find and understand the fundamental causes of our cybersecurity challenges. For this, we need to discover the first principles of cybersecurity.

Appendix A: An SIS Model of Epidemiological Outbreaks, Thresholds, and Their Relation to Cybersecurity

As discussed in Chapter 6, my understanding of how graph spectral analysis can help us reduce the complexity of a system is largely informed by the work of Chakrabarti, et. al. Below I reproduce Chakrabarti's results in more detail than in his original work. I am doing this mainly to support the notion that his conclusions are defensible, but also partially to document for myself and others the simplifications and assumptions that Chakrabarti made in his papers. As an engineer, my experience with formal proofs is somewhat limited. As a result, I was interested in recreating Chakrabarti's work to prove to myself that I understood the concepts. It is my desire that this augmented explanation of his work will help others interested in applying spectral analysis, specifically the concept of spectral radius, to cybersecurity.

Chakrabarti starts off with a model of the epidemiological process of viral outbreaks. He uses a susceptible, infected, and susceptible (SIS) model for his purposes. In general, the idea is that in a population of people during a viral outbreak, people are either susceptible to the virus or infected with the virus. The model, as defined by Chakrabarti, is a discrete time model that is based on the birth rate (β) and death rate (δ) of the virus; it uses probabilities to predict the likelihood of those susceptible being infected, or those infected being cured.

The model is formalized in two parts. First, he defines the probability that a node *will not* be infected at time *t* based on its neighbors from the previous time step (*t-1*). If the node in question is *i*, and the neighbor is *j*, *i* does not get infected from *j* if *j* is not infected (probability 1-$p_{j,t-1}$), or *j* is infected but does not infect *i* (i.e. the probability *j* is infected, $p_{j,t-1}$, multiplied by the probability the virus is not passed, *1-β*). The total probability that *i* remains uninfected at *t* $(\xi_{i,t})$, is the product of the probabilities from each neighbor of *i*.

$$\xi_{i,t} = \prod_{j\ as\ a\ neighbor\ of\ i} \left(p_{j,t-1}(1-\beta) + (1-p_{j,t-1})\right)$$

This can be reduced to:

$$\xi_{i,t} = \prod_{j\ as\ a\ neighbor\ of\ i} (1-\beta * p_{j,t-1})$$

Because all terms are positive, we can reduce the above equation to an inequality that simplifies our analysis while ensuring that we are providing conservative approximation of the probability that *i* is not infected by any neighbor *j*.

$$\xi_{i,t} \geq 1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}$$

By making this simplification, we are asserting that the probability that *i* is not infected from *j* is at least $\xi_{i,t}$ and may be even greater.

The second part of the model of the probability of *i* not being infected at time *t* considers the probability that *i* is not already infected and the probability that *i* was infected at *t*-1, but cured by time *t*. With these additions, the model of node *i* not being infected at time *t* is written:

$$1 - p_{i,t} = (1 - p_{i,t-1})\xi_{i,t} + \delta p_{i,t-1}\xi_{i,t} \quad where\ i = 1 \dots N$$

Since $p_{i,t-1}$ is the probability that node *i* is infected at time *t*-1, 1-$p_{i,t-1}$ is the probability that node *i* is not infected at time *t*-1. Substituting the expression for $\xi_{i,t}$ we get:

$$1 - p_{i,t} \geq (1 - p_{i,t-1})\left(1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}\right) + \delta p_{i,t-1}\left(1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}\right) \quad where\ i = 1 \dots N$$

$$1 - p_{i,t} \geq \left((1 - p_{i,t-1}) + \delta p_{i,t-1}\right)\left(1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}\right)$$

$$1 - p_{i,t} \geq (1 - p_{i,t-1} + \delta p_{i,t-1})\left(1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}\right)$$

$$1 - p_{i,t} \geq (1 - (1 - \delta)p_{i,t-1})\left(1 - \beta * \sum_{j\ as\ a\ neighbor\ of\ i} p_{j,t-1}\right)$$

Since the summation operator is summing elements *j* that are neighbors of elements *i*, we can write the summation using the adjacency matrix of *i*, and sum from 1 to *N*.

$$1 - p_{i,t} \geq (1 - (1 - \delta)p_{i,t-1})\left(1 - \beta * \sum_{j=1}^{N} \boldsymbol{A}_{i,j} * p_{j,t-1}\right)$$

Expanding the inequality, we get:

$$1 - p_{i,t} \geq \left(1 - (1 - \delta)p_{i,t-1}\right) - \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right) + \left(\beta(1 - \delta)p_{i,t-1} * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)$$

We can further simplify the inequality and make it more conservative by noting that the final term on the right side of the inequality will always be a positive number. We can state this because the variables are all probabilities and hence have a value of at most 1. Therefore, even though the term has a subtraction operation, we are assured to never get a result less than zero. Considering this, we can eliminate this final term to further our conservative approximation of the probability that node *i* is not infected. Ignoring this term has the added benefit of simplifying the inequality for further analysis. We are then left with the inequality:

$$1 - p_{i,t} \geq \left(1 - (1 - \delta)p_{i,t-1}\right) - \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)$$

Chakrabarti goes on to use this model to prove that a system defined by this model is subject to an epidemic outbreak if the ratio of the birthrate to the deathrate of the virus surpasses a threshold. Chakrabarti proves that this threshold is related to the largest eigenvalue of the graph that describes the network. I detail his proof below.

We assume that a threshold, τ, for the epidemic spread of a virus is related to the ratio of the birthrate, β, to the deathrate, δ, of the virus.

$$\tau = \frac{\beta}{\delta}$$

We posit that the threshold is equal to the reciprocal of the largest eigenvalue of the adjacency matrix of the graph representing the network, $\lambda_{1,A(G)}$.

$$\tau = \frac{1}{\lambda_{1,A(G)}}$$

Chakrabarti proves this by starting with the model of the system. From here, he argues that the system must be asymptotically stable as the infection rate approaches zero. In other words, the system must tend to zero as the infection rate gets closer to zero, and not have any surprise stochastic behavior that drives the system infection rate higher at some point as the system approaches zero. To prove this asymptotic stability, he cites a proof from Hirsch, Smale, and Devaney [73] that demonstrates a system is asymptotically stable at *P*=0 if the magnitude of the eigenvalues of the system are less than 1. To calculate these eigenvalues, Hirsch, et. al. uses the equation:

$$\left[\nabla f(\vec{0})\right]_{i,j} = \left.\frac{\partial f_i}{\partial p_j}\right|_{\vec{P}=\vec{0}}$$

The idea is to take the partial derivative of the function, set the remaining *p* terms to zero, set the equation to less than or equal to 1, and solve. We will see that with the simplifications that we made earlier the *p* terms drop out because the simplified equation is first order.

We apply this to the function we defined earlier:

$$1 - p_{i,t} \geq \left(1 - (1-\delta)p_{i,t-1}\right) - \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)$$

Arranging the inequality to fit the equation:

$$p_{i,t} \leq 1 - \left(1 - (1-\delta)p_{i,t-1}\right) + \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)$$

which we will write using a vector $\vec{P}$ for p_i that includes all nodes i

$$\vec{P}_t \leq f\left(\vec{P}_{t-1}\right),$$

where

$$f_i\left(\vec{P}_{t-1}\right) \leq 1 - \left(1 - (1-\delta)p_{i,t-1}\right) + \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)$$

Performing the partial derivative on the function $f_i\left(\vec{P}_{t-1}\right)$ we get:

$$\left.\frac{\partial f_i}{\partial p_j}\right|_{\vec{P}=\vec{0}} \leq \left.\frac{\partial\left(1 - \left(1 - (1-\delta)p_{i,t-1}\right) + \left(\beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\right)\right)}{\partial p_j}\right|_{\vec{P}=\vec{0}}$$

$$\leq \left.\frac{\partial\big(1 - 1 + (1-\delta)p_{i,t-1} + \beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\big)}{\partial p_j}\right|_{\vec{P}=\vec{0}}$$

$$\leq \left.\frac{\partial\Big((1-\delta)p_{i,t-1} + \beta * \sum_{j=1}^{N} A_{i,j} * p_{j,t-1}\Big)}{\partial p_j}\right|_{\vec{P}=\vec{0}}$$

If *i≠j*, the result is:

$$\leq \beta * A_{i,j}$$

If *i=j*, the result is:

$$\leq \left.\frac{\partial\Big((1-\delta)p_{i,t-1} + \beta * \sum_{j=1}^{N} A_{i,i} * p_{i,t-1}\Big)}{\partial p_i}\right|_{\vec{P}=\vec{0}}$$

$$\leq (1-\delta) + \beta * A_{i,i}$$

Since $A_{i,i}$ are the diagonal terms of the graph, they are all zero, hence:

$$\leq (1-\delta)$$

We will write the inequality as an equality assuming the most conservative case:

$$\big[\nabla f(\vec{0})\big]_{i,j} = \begin{cases} \beta * A_{i,j}\ for\ j\ \neq i \\ (1-\delta)\ for\ j = i \end{cases}$$

We will reform this equation as:

$$\mathbf{S} = \nabla f(\vec{0}) = \ \beta\mathbf{A} + (1-\delta)\mathbf{I}$$

Chakrabarti goes on to prove that the eigenvalues of **S** take the form:

$$\lambda_{i,S} = 1 - \delta + \beta\lambda_{i,A} \quad \forall i$$

Using the stability criteria stated above the system is asymptotically stable if:

$$|\lambda_{i,S}| < 1 \quad \forall i$$

Chakrabarti goes on to explain that because **A** is a real symmetric matrix and the graph is undirected, its eigenvalues are real, hence the eigenvalues of **S** are real. Additionally, the largest eigenvalue is a positive real number and has the largest magnitude of all eigenvalues. This allows us to state:

$$\lambda_{1,S} = |\lambda_{1,S}| \geq |\lambda_{i,S}| \quad \forall i \quad where\ \lambda_{1,S}\ is\ the\ largest\ eigenvalue\ of\ \mathbf{S}$$

Thus:

$$|\lambda_{i,S}| < 1 \quad \rightarrow \lambda_{1,S} < 1$$

Substituting for $\lambda_{1,S}$:

$$1 - \delta + \beta\lambda_{1,A} < 1$$

$$-\delta + \beta\lambda_{1,A} < 0$$

$$\beta\lambda_{1,A} < \delta$$

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{1,A}}$$

Hence, the stability of the network is maintained when the reciprocal of the largest eigenvalue is larger than the ratio of the birthrate to the deathrate.

References

[1] Center for Strategic and International Studies, "Significant Cyber Incidents," CSIS, 6 2021. [Online]. Available: https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents. [Accessed 7 7 2021].

[2] N. Perlroth and C. Krauss, "New York Times," New York Times, 15 3 2018. [Online]. Available: https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html. [Accessed 7 7 2021].

[3] F. Robles and N. Perlroth, "New York Times," New York Times, 8 2 2021. [Online]. Available: https://www.nytimes.com/2021/02/08/us/oldsmar-florida-water-supply-hack.html. [Accessed 7 7 2021].

[4] K. Zetter, "Wired," Conde Nast, 20 1 2016. [Online]. Available: https://www.wired.com/2016/01/everything-we-know-about-ukraines-power-plant-hack/. [Accessed 7 7 2021].

[5] C. E. Landwehr, "Cybersecurity: From engineering to science," *The Next Wave,* vol. 19, no. 2, pp. 2-4, 2012.

[6] National Security Agency, "Science of Security," [Online]. Available: https://www.nsa.gov/what-we-do/research/science-of-security/. [Accessed 21 08 2021].

[7] Willis H. Ware, "Security Controls for Computer Systems: Report of Defense Science Board Task Force on Computer Security," Rand Corporation, Santa Monica, 1969.

[8] S. Xu, "Cybersecurity Dynamics: A Foundation for the Science of Cyber Security," in *Proactive and Dynamic Network Defense*, Springer International Publishing, 2019, pp. 1-31.

[9] J.-H. Cho and S. e. a. Xu, "STRAM: Measuring the Trustworthiness of Computer-based Systems," *ACM Computing Surveys,* vol. 128, 2019.

[10] J. Coleman, The Foundations of Social Theory, Cambridge: Harvard Unversity Press, 1994.

[11] Aristotle and J. Barnes, Posterior Analytics, Oxford: Clarendon Press, 1976.

[12] M. Gasser-Wingate, "Aristotle of Induction and First Principles," *Philosophers' Imprint,* vol. 16, no. 4, 2016.

[13] R. Descartes and J. Veitch, The Principles of Philosophy, Coppell, 2019.

[14] J. P. Anderson, "Computer Security Technology Planning Study," USAF, Bedford, 1973.

[15] Bell, David E; LaPadula, Len;, "SECURE COMPUTER SYSTEMS: MATHEMATICAL FOUNDATIONS," MITRE, Bedford, 1973.

[16] D. E. Bell, Interviewee, *An Interview with David Elliott Bell.* [Interview]. 12 September 2012.

[17] J. Saltzer and M. D. Schroeder, "The Protection of Information in Computer Systems," *Proceedings of the IEEE,* vol. 63, no. 9, pp. 1278-1308, 1975.

[18] J. R. Yost, "The Origin and Early History of the Computer Security Software Products Industry," *IEEE Annals of the History of Computing,* pp. 46-58, 2015.

[19] S. Sammonas, "THE CIA STRIKES BACK: REDEFINING CONFIDENTIALITY, INTEGRITY AND AVAILABILITY IN SECURITY," *Journal of Information System Security,* vol. 10, no. 3, pp. 21-45, 2014.

[20] International Standards Organization, *Information technology — Security techniques — Information security management systems — Overview and vocabulary,* Geneva: International Standards Organization, 2018.

[21] B. Lundgren and N. Moller, "Defining Information Security," *Science and Engineering Ethics,* vol. 25, no. 2, pp. 419-441, 2019.

[22] Mason, C.R., The Art and Science of Protective Relaying, Schenectady: Wiley, 1956.

[23] K. Godel and B. Meltzer, On Formally Undecidable Propositions of Principia Mathematica and Related Systems, Mineola: Dover, 1992.

[24] Palo Alto Networks, "What is Zero Trust?," Palo Alto Networks, [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/what-is-a-zero-trust-architecture. [Accessed 23 09 2020].

[25] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," *Zeitschrift für Physik,* vol. 43, no. 3, pp. 172-198, 1927.

[26] Standford University, "Stanford Encyclopedia of Philosophy," 2 February 2008. [Online]. Available: https://plato.stanford.edu/entries/certainty/. [Accessed 22 March 2021].

[27] L. Brillouin, "Maxwell's Demon Cannot Operate: Information and Entropy," *Journal of Applied Physics,* vol. 22, no. 3, pp. 334-337, 1950.

[28] L. Szilard, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings," *Zeitschrift fur Physik,* vol. 53, pp. 840-856, 1929.

[29] P. Rodd, "Some Comments on Entropy and Information," *American Journal of Physics,* vol. 32, no. 5, pp. 333-335, 1964.

[30] M. Tribus, "Methods of Statistical Inference," in *Thermostatics and Thermodynamics*, Dartmouth, D. Van Nostrand Company, Inc, 1961, pp. 29-66.

[31] K. Jones, "Trust as an Affective Attitude," *Ethics,* vol. 107, no. 1, pp. 4-25, 1996.

[32] D. Gambetta, Trust: Making and Breaking Cooperative Relations, Oxford: electronic edition, 2000.

[33] V. McGreer, "Truat, Hope, and Empowerment," *Australasian Journal of Philosophy ,* vol. 86, no. 2, pp. 237-254, 2008.

[34] W. T. Hardwood, "The Logic of Trust," electronic copy, York, 2012.

[35] N. Luhmann, M. King and C. Morgner, Trust and Power, Cambridge: Polity Press, 2017.

[36] Wikipedia, "MD5," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/MD5. [Accessed 27 09 2020].

[37] Office of the Directorate of National Intelligence, "National Security Adjudicative Guidelines," Office of the Directorate of National Intelligence, Washington, D.C., 2017.

[38] A. K. Fedorov, E. O. Kiktenko and A. I. Lvovsky, "Quantum computers put blockchain security at risk," *Nature,* 19 11 2019.

[39] A. Shenfield, D. Day and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," *ICT Express,* vol. 4, no. 2, pp. 95-99, 2018.

[40] North American Transmission Forum, "Supply Chain Cyber Security Industry Coordination," [Online]. Available: https://www.natf.net/industry-initiatives/supply-chain-industry-coordination. [Accessed 28 09 2020].

[41] Office of the President of the United States of America, *Executive Order 13920 Securing the United States Bulk Power System,* Distric of Columbia: National Archives, 2020.

[42] The Office of the President of the United States of America, "Executive Order 14028," 12 05 2021. [Online]. Available: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/. [Accessed 21 09 2021].

[43] S. P. Marsh, "Formalising Trust as a Computational Concept," Stirling, 1994.

[44] H. A. Simon, "A behavioural model of rational choice," *Quarterly Journal of Economics,* vol. 69, pp. 99-118, 1955.

[45] R. Feldman, J. Forrest and B. Happ, "Self-presentation and verbal deception: Do self-presenters lie more?," *Basic and Applied Social Psychology,* vol. 24, pp. 163-170, 2002.

[46] S. Lloyd, "Complexity," [Online]. Available: https://web.mit.edu/esd.83/www/notebook/Complexity.PDF. [Accessed 12 10 2020].

[47] F. Heylighen and e. al., "Complexity and Philosophy," in *Complexity, Science and Society*, Brussels, Radcliffe, 2005.

[48] J. Ladyman, J. Lambert and K. Wiesner, "What is a Complex System?," *European Journal for Philosophy of Science,* vol. 3, no. 1, 2013.

[49] C. R. Shalizi and J. P. Crutchfield, "Computational Mechanics: Pattern and Prediction, Structure, and Simplicity," 2008.

[50] H. Simon, "The Architecture of Complexity," *Proceedings of the American Philosophical Society,* vol. 106, no. 6, pp. 467-482, 1962.

[51] J. Flack, "Life's Information Hierarchy," in *Worlds Hidden in Plain Sight*, Santa Fe, The Santa Fe Institute Press, 2019, pp. 201-225.

[52] NIST, "Computer Security Resource Center," NIST, 5 8 2015. [Online]. Available: https://csrc.nist.gov/Projects/Hash-Functions/NIST-Policy-on-Hash-Functions. [Accessed 7 7 2021].

[53] Schweitzer Engineering Laboratories, "Secure System Overview," Schweitzer Engineering Laboratories, [Online]. Available: https://selinc.com/solutions/sfci/secure-system-overview/. [Accessed 8 11 2020].

[54] T. O'Conner, "Emergent Properties," Stanford University, Fall 2020. [Online]. Available: https://plato.stanford.edu/entries/properties-emergent/. [Accessed 8 11 2020].

[55] R. K. Standish, "On Complexity and Emergence," *arVix,* 2001.

[56] R. C. Armstrong, J. R. Mayo and F. Siebenlist, "Complexity Science Challenges in Cyber Security," Sandia National Laboratories, Livermore, 2009.

[57] P. W. Anderson, "More is Different," *Science,* vol. 177, no. 4047, pp. 393-396, 1972.

[58] M. Christen and L. R. Franklin, "The Concept of Emergence in Complexity Science: Finding Coherence between Theory and Practice," 2004.

[59] S. Xu, "Emergent Behavior in Cybersecurity," in *HOTSOS*, Raleigh, 2014.

[60] D. CHAKRABARTI, Y. WANG, C. WANG, J. LESKOVEC and C. FALOUTSOS, "Epidemic Thresholds in Real Networks," *ACM Transactions on Information and System Security,* vol. 10, no. 4, 2008.

[61] M. Li and P. Vitiyani, "A Brief Introduction," in *An Introduction to Kolmogorov Complexity and Its Applications* , New York, Springer, 2019, pp. 1-7.

[62] N. Immerman, Descriptive Complexity, New York: Springer, 1999.

[63] J. Reason, Human Error, Cambridge: Cambridge University Press, 1990.

[64] J. Gribbin, Deep Simplicity Bringing Order to Chaos and Complexity, New York: Random House, 2004.

[65] R. Pastor-Satorras, C. Castellano and P. V. Mieghem, "Epidemic processes in complex networks," *REVIEWS OF MODERN PHYSICS,* vol. 87, 2015.

[66] D. Cvetkovic, P. Rowlinson and S. Simic, An Introduction to the Theory of Graph Spectra, London: Cambridge University Press, 2010.

[67] P. V. Mieghem and e. al., "Virus Spread in Networks," *IEEE/ACM Transactions on Networking,* vol. 17, no. 1, 2009.

[68] P. V. Mieghem, D. Stevanovic, F. Kuipers, C. Li, R. v. d. Bovenkamp, D. Liu and H. Wang, "Decreasing the spectral radius of a graph by link removals," *PHYSICAL REVIEW E,* vol. 84, no. 1, 2011.

[69] M. Xu, G. Da and S. Xu, "Cyber Epidemic Models with Dependences," *Internet Mathematics,* vol. 11, no. 1, pp. 62-92, 28 03 2015.

[70] F. K. Basha and e. al., "Implementation of Reliable High-Speed Islanding Detection, Zone Interlocking, and Source Selection Schemes Using Smart Algorithms," in *Industrial & Commercial Power Systems Technical Conference*, Calagary, 2015.

[71] MITRE Corporation, "Science of Cyber Secuity," The MITRE Corporation, McLean, 2010.

[72] J. Erickson, "NP-Hardness," in *Algorithms*, Urbana, self-published, 2019, p. 382.

[73] M. W. Hirsch, S. Smale and R. L. Devaney, Differential Equations, Dynamical Systems, and Linear Algebra, Elsevier Science, 1974.