The
Information Security Practice Principles

Foundational Whitepaper

Version 0.9
May 2017

CENTER FOR APPLIED
CYBERSECURITY RESEARCH
INDIANA UNIVERSITY
Pervasive Technology Institute

Authors & Acknowledgements

This work was authored by Craig Jackson, Scott Russell, and Susan Sons. Craig Jackson (scjackso@indiana.edu) is project lead.

The authors thank Von Welch and the Indiana University Center for Applied Cybersecurity Research (CACR) for supporting this effort. The authors thank their colleagues at IU and CACR, in addition to LTC Ernest Wong and Ms. Stephanie Pell of the Army Cyber Institute for their invaluable input. Special thanks to Dr. Rob Templeman of Naval Surface Warfare Center Crane Division for his insight and comments.

About CACR

The Indiana University Center for Applied Cybersecurity Research (CACR, cacr.iu.edu) is distinctive in addressing cybersecurity from a comprehensive, multidisciplinary, hands-on perspective. The Center draws on Indiana University's scholarly expertise in computer science, informatics, accounting and information systems, criminal justice, law, regulatory compliance, organizational behavior, and public policy. Founded by now-IU President Michael McRobbie in 2003, CACR's recent work includes research awards from DHS, DOE, NSF, and collaborations with NSWC Crane Division, the Indiana National Guard, the City of Chicago, and Indiana's legal community. CACR leads the National Science Foundation's Cybersecurity Center of Excellence. Under CACR's coordination, NSA and DHS designated IU a National Center of Academic Excellence in Cyber Defense Research and Information Assurance/Cybersecurity Education. CACR is part of Indiana University's Pervasive Technology Institute.

Citing this Work

This work is made available under the terms of the Creative Commons Attribution 4.0 International license. For more details, please visit the following URL: https://creativecommons.org/licenses/by/4.0/.

Table of Contents

About the ISPPs

We believe high-level principles underlie a great deal of existing information security[1] thinking and practice, but that they have remained generally under-researched and unarticulated in favor of technical documents that are highly detailed and highly prescriptive, such as the NIST Risk Management Framework, CIS Critical Security Controls, ISO standards, or the HIPAA security rule. These documents may be loaded with great advice, but they are difficult to understand without the benefit of significant prior training, and do little to help someone learn to "think like a security practitioner" or to address novel, emergent situations. The Information Security Practice Principles seek to bridge this gap, providing a foundational mental model for information security problem-solving. The Principles can be used to teach new or non-practitioners, such as students and executives, about doing information security; they can help practitioners make decisions in novel situations, where an established best practice may not exist; and they can add validity and salience to existing, more-detailed statements of best practice.

> Principle (n.):
> A general law or rule adopted or professed as a guide to action; a settled ground or basis of conduct or practice; a fundamental motive or reason for action, esp. one consciously recognized and followed.
>
> *Oxford English Dictionary, online*

The principles described in this paper are "guide[s] to action." Principles like these do not describe an end-state where perfect (or reasonable or acceptable) security is achieved. Unlike, for example, the CIA triad[2] or the concept of *resilience,*[3] they do not describe security goals or objectives.[4] Rather, the Principles guide decision-making.

Motivation

Information security practitioners, as well as the people who rely on and supervise them, need a foundational mental model for decision-making in complex situations where there is no single "right answer." Information security is not a solved problem. It is a challenging, multi-disciplinary domain where defenders are faced with, must diagnose, and must

[1] We use "information security" and "cybersecurity" interchangeably. We have chosen to use the former in this work to emphasize that our scope includes computing environments (whether networked or not) as well as information, information systems, and information system elements (including humans) that exist primarily or entirely in the physical domains.

[2] The Confidentiality-Integrity-Availability triad is the set of security objectives for information and information systems outlined in FISMA, 44 U.S.C. § 3541, et seq.

[3] *See*, Presidential Policy Directive 21 (PPD-21): Critical Infrastructure Resilience and Security. https://www.whitehouse.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil.

[4] *See also*, Donn Parker, "Toward a New Framework of Information Security," http://www.computersecurityhandbook.com/csh4/chapter5.html (expanding CIA to include "utility," "authenticity," and "possession," referred to as the "Parkerian Hexad").

respond to a wide range of threats, as well as to the demands of a complex social, psychological, and economic environment. Much of the information security canon is highly detailed, narrowly applicable, and highly prescriptive. Resources of this kind can be valuable, particularly where what is reasonable or “best practice” is well-defined, but they do not on their own help anyone learn to “think like an information security practitioner.”

> Our purpose is to identify the underlying and invariant principles that inform information security generally; those which have driven and guided information security decision-makers across technologies, sectors, and epochs.

The very best information security professionals are like health care professionals, lawyers, and military commanders. They do much more than implement compliance checklists or set up firewalls: they think critically and use judgment to make decisions and offer guidance. They apply their experience and expertise to the full scale of cyber problems, from system design to developing and implementing cybersecurity programs addressed to an entire mission or campaign.

We need more of these cyber samurai, and that means maturing the information security community and how we educate and train. There are true masters of information security, but we believe that excellence in this field leans heavily on master-apprentice relationships, trial-and-error experience, and the mimetic transfer of knowledge. These represent very powerful ways to learn, but they don’t necessarily scale or produce quick results. The ISPPs can be a cornerstone of information security education, helping new practitioners build a very deep and very broad insight into what information security is all about, not unlike the Fair Information Practice Principles[5] for privacy professionals, or the Model Rules of Professional Conduct[6] for lawyers.

Moreover, we need more people up and down the chain of command and in other specialties to be able to engage in and understand the information security dialogue at a base level. The intended audience for this work is anyone making security decisions at any level of an organization: current practitioners, future practitioners, technologists whose decisions interact with or impact security, managers, and stakeholders.

Methodology

To uncover and set out the Principles, we began with a review of prior work, assessing the

[5] For a multinational analogue to the Fair Information Practice Principles (FIPPs), *see* Organization for Economic Cooperation and Development Guidelines on the Protection of Privacy and the Transborder flow of Personal Information, which established eight “privacy principles” that have served as the foundation for data protection regimes across the world, most notably in the European Union. This analogue is important because the OECD also created Guidelines on the Security of Information Systems, discussed below, which serve largely as a point of contrast to the Information Security Practice Principles.

[6] “Model Rules of Professional Conduct,” American Bar Association, *available at* https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_table_of_contents.html.

need for and feasibility of carrying forward with this project. We then conducted a broad search and review within information security and across related fields to find evidence of the Principles at work.[7] Finally, and in an iterative refinement process against our broad review, we applied selection and tailoring criteria to the results of our search to build this report and related artifacts. Below, we describe our methodology and criteria in detail.

Step 1: Prior Work.

We began with a literature review. Specifically, we sought to identify past efforts in the information security community to collect and explicate the authoritative set of information security principles, in the sense of the OED definition we adopted. The results of this literature review brought in a wide-range of sources from a diverse set of backgrounds, including academic journal articles, industry publications, governmental standards and policy documents, blog posts, course materials,[8] expert opinions, and text books.[9] In total, we collected over 50 sources. Notable among these were Saltzer and Schroeder's renowned "The Protection of Information in Computer Systems," as well as the National Institute of Science and Technology's "Generally Accepted Principles and Practices of Information Security,"[10] the OECD Guidelines on the Security of Information Systems and Networks,[11] and the Information Systems Security Association's "Generally Accepted Information Security Principles."[12,13] While this is by no means an exhaustive list, we identify these sources in particular for their proximity to our stated goal, and for providing particularly authoritative, innovative, or thorough perspectives on the question of information security principles.

However, even among these preferred sources, none satisfied our desired goal for a complete, universal, scalable, and actionable set of guiding principles. A recurrent shortcoming was that these sources utilized weak or loose definitions of "principle," and therefore struggled to distinguish true principles from the "best of best practices" or from vague normative goals. Indeed, the majority of sources appeared to be attempts to compile or prioritize established best practices. Moreover, these sources were frequently opaque in terms of selection criteria for which practices should be included, how to justify inclusion,

[7] Our search was limited to unclassified materials.
[8] *See, e.g*., Michael Clarkson, "Principles of Security," Cornell, http://www.cs.cornell.edu/courses/cs5430/2015sp/notes/principles.php.
[9] Although there is a prominent information security textbook titled "Principles of Information Security," this usage does not reflect our intended target with the word "principle," and so this source is better considered under the information security archeology section, discussed below.
[10] Marrianne Swanson & Barbara Guttman, "Generally Accepted Principle and Practices for Securing Information Technology Systems," NIST SP 800-14, http://csrc.nist.gov/publications/nistpubs/800-14/800-14.pdf.
[11] "OECD Guidelines for the Security of Information Systems and Networks," OECD, (July 25, 2002), http://www.oecd.org/sti/ieconomy/15582260.pdf (hereinafter "OECD Guidelines").
[12] Generally Accepted Information Security Principles, ISSA, http://all.net/books/standards/GAISP-v30.pdf.
[13] Other notable sources include: Gary McGraw, "Thirteen principles to ensure enterpise system security," SearchSecurity, (Jan. 2013) http://searchsecurity.techtarget.com/opinion/Thirteen-principles-to-ensure-enterprise-system-security. *See also* "The 7 Basic Principles of IT Security," Technopedia, (Dec. 5, 2014), https://www.techopedia.com/2/27825/security/the-basic-principles-of-it-security.

and why the selected practices furthered security in a broader sense.[14] On the other hand, those sources that identified more normative "principles" often proved too vague to guide security decision-making at a practical level, instead offering overarching goals that were not directly related to security.[15]

More problematic still was that these sources often proved to be incomplete or narrowly scoped, focusing on one aspect of information security, such as risk management, to the exclusive of others.

More broadly, the sources we identified consistently failed to justify *why* the principles they identified were fundamental to information security. Although these sources provided effective, important, and illustrative information, they never connected the dots as to why and how the specific practices or normative goals furthered information security at a baseline level.

Therefore, based on the conceptual gap identified in the literature review, we felt confident that the ISPPs could be a meaningful contribution to the information security community. Using the shortcomings we identified as a baseline and a richer sense of our selection criteria (see Step 3), we dug deeper into the web of sources and wisdom impacting information security.

Step 2: Information Security Archaeology.

To ensure that the ISPPs identified the scope and breadth of information security, we set out on a deep dive into the range of sources, fields, and disciplines which inform information security. Our sources ranged from information security frameworks and governance documents, to compilations of best practices, to open-source standards and security community knowledge. We also searched beyond information security sources, drawing from related fields, with particular attention to fields focused on protection, adversity, secrecy, decision-making, and trust. These included military doctrine, information theory, complexity theory, risk management, disaster recovery, engineering assurance, law, medicine, ecological psychology, and intelligence.[16] With this informational backdrop, we sought to identify, categorize, abstract, and consolidate these wide-ranging sources into a short series of concrete, discrete, and scaleable principles. We then reanalyzed the existing source material to ensure that each of the preexisting efforts was fully encompassed by our list, and that all commonly cited principles are either directly covered or derivable from the ISPPs. By utilizing both a bottom-up and top-down approach, our goal was ensure that our list was comprehensive and inclusive, while still providing the opportunity for the ISPPs to deviate from the established norms when we felt that the underlying principle for a

[14] This is not to infer that these attempts to compile information security best practices are not valuable resources. However, it is important to distinguish these compilations from the underlying principles that articulate why those best practices are valuable and how they fit into the broader picture of information security.

[15] For instance, the OECD Guidelines identify "Democracy" as a principle, stating that "The security of information systems and networks should be compatible with essential values of a democratic society." While undoubtedly a noble and important goal, we found this and principles like it to be too vague to guide information security decision making at a practical level, and therefore did not meet our criteria for a "principle."

[16] The specific sources from these fields will be referenced in the accompanying essays.

given practice has not been appropriately identified.

Step 3: Selection Criteria and Curation.

To make sense of our broad, deep search, we employed the following criteria to identify and describe the principles we found.

For each derived principle, we required the following qualities:

1. Grounded in prior work. The principle had to be clearly grounded in prior work in information security as well as as have some historical grounding or corollary in a related field.
2. Guiding of action. The principle had to be expressible as a guide to action, meeting the definition of principle discussed earlier. This meant that the principle had to do more than express an attitude, outcome, or a factual truth. Outcomes, objectives, or goals like resilience, confidentiality, integrity, or availability are certainly worthy, but do not in and of themselves directly guide action. Truisms like "you are only as secure as your weakest link" might suggest the existence of a principle, but themselves are insufficiently imperative or guiding to make the cut. Moreover, each principle had to have a useful degree of specificity. For instance, were we searching for moral principles rather than information security principles, "Be a good person" and "Be nice" would have been insufficient, whereas the Golden Rule ("Do unto others as you would have done to you") is sufficiently specific. The Golden Rule actually instructs us on how to get there.
3. Directly related to outcomes. Each principle had to have a proximate causal[17] relationship to desired security outcomes and support security assessment. Fault Tolerance and its definition only work if we can clearly trace a causal relationship between "anticipat[ing] and address[ing] the potential compromise and failure of system elements and security controls" and resilient systems or critical data that remains available in a contested environment. This criterion was important to us as archaeologists and curators because so many phenomena have some causal connection to positive (or negative) outcomes. Moreover, this criterion ensured that the principles can be used not only to plan for, but also to explain security successes and security failures after the fact.
4. Generally applicable through time and space. Each principle had to be generally applicable through time and space. This meant that principles had to be technology neutral and applicable to information security both inside and outside of computing contexts. Additionally, each principle had to have a sufficient historical grounding as to be timeless and to give us confidence in the principle's invariance. Each principle has to be lifecycle neutral, from system design, control design, and supply chain evaluation through to *in situ*, in the moment decision-making. Finally, each principle had to be scalable in application, from system elements, to systems, to systems of systems, to, often most importantly, missions.

To create an impactful set of principles, we employed the following criteria:

1. Sufficiently inclusive. The set had to be sufficiently inclusive to encompass the full

[17] Black's Law Dictionary offers as the primary definition of proximate cause: "That which, in a natural and continuous sequence, unbroken by any efficient intervening cause, produces injury, and without which the result would not have occurred."

scope of information security practice. For us, information security is a complex, multidisciplinary affair, and people we think of as practitioners operate at different levels and with a variety of expertises. Our principles are crafted to be meaningful across this wide range of disciplines and expertises.

2. Internally consistent. Each principle had to be expressible with a similar level of abstraction and with consistent use of core terminology. To work as a set, the principles must read like a set, even if each one is not equally applicable to each real world problem, and even if they come into tension when applied in practice.
3. Useful in combination. The principles had to be usable in combination to explain more specific practices, controls, and events. Just as we required a strong causal, explanatory connection between individual principles and security outcomes, we required a strong explanatory connection between combinations of principles and real world phenomena. For instance, "least privilege" is often referred to as a principle of information security. We knew that least privilege either had to be one of the principles, or be explainable as a combination[18] of the principles.
4. 7 +/- 2. We aimed for a set of principles that is short enough to remember and readily bring to mind, but not so short as to obscure the guidance.

The remainder of this paper consists of our Operational Definitions and the seven essays introducing the seven principles. There are many terms to define in this paper, but a few form the core conceptual, ontological glue that holds the principles together. Each essay features the Principle's definition, scope, how it manifests in the real world, how it interacts with other principles, and its impact on information security. Each "Grounding" section dives deeper into the background research and conceptual basis for the principle. Each "Strategies & Challenges" section explores ways to implement the principle and work through the inevitable challenges of doing information security well.

[18] For a popular analogy, these combinations of principles can be thought of as information security "cocktails." For example, least privilege could be: two parts Compartmentation, two parts Minimization, and one part Proportionality. (And when made really well, a dash of Rigor.)

Operational Definitions

In order to describe these high level principles with some consistency, we've found it necessary to set operational definitions for a limited set of terms. Together, these concepts provide an ontology[19] with which security practitioners can describe the full range of information security scenarios. These definitions are pragmatic: they are not meant to describe reality as it actually is, they are designed to be a minimal, cohesive set of mental tools for applying the principles to security situations. The terms and definitions are inspired by complexity science,[20] ecological psychology,[21] quantum physics,[22] the Oxford English Dictionary, and (of course) information security.[23]

Agency Terms

1. **Actor:** Entities capable of perception, decision-making, and intelligent action. This term emcompasses humans, organizations, and (maybe) strong artificial intelligence.
2. **Security Practitioner**: A human actor responsible for the formulation, implementation, evaluation, or selection of security strategies and controls, and the promotion or design of secure systems architecture.

Spatiotemporal Terms

3. **Environment:** The totality of actual and potential phenomena that may impact a mission. Environments include actors, risks, natural phenomena, systems, and system elements.
4. **Information System ("System"):** An identifiable set of interconnected elements organized for the collection, processing, maintenance, use, sharing, dissemination, or disposition of information. Systems are defined pragmatically and can range from entire organizations to individual software applications to individual microchips.
5. **Information System Element ("System Element" or "Element"):** A discrete, identifiable part of a system. Information system elements may include information itself (whether analogue or digital), human actors, devices, networks, software, and other systems.
6. **Risk:** An uncertainty which may negatively impact the mission.[24]

Intentionality Terms

7. **Mission:** The overarching goal or goals that an information security strategy enables. The mission provides the primary lens through which security practitioners make sense of the environment and identify risks.
8. **Information Security Strategy ("Security Strategy")**: A plan for successful mission support based on coordinated application of information security controls.
9. **Information Security Control ("Security Control" or "Control")**: An administrative, technical, or physical safeguard or countermeasure operating within the environment to address a risk. Security controls include detective, preventative, responsive, and corrective controls.

[19] We use "ontology" here in the information science sense, *i.e.*, "explicit formal specifications of the terms in the domain and relations among them," *see*, Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition* 5: 199-220.

[20] *See, e.g.*, Burns et al., "Organizational Information Security as a Complex Adaptive System: insights from three agent-based models," Inf Sys Front (2015). doi:10.1007/s10796-015-9608-8.

[21] *See, e.g.*, J. Gibson "The Theory of Affordances. In Perceiving, Acting, and Knowing," Robert Shaw and John Bransford, ets. (1977). ISBN 0-470-99014-7.

[22] *See, e.g.,* Gösta Ekspong, "The Dual Nature of Light as Reflected in the Nobel Archives," (Dec. 2, 1999), http://www.nobelprize.org/nobel_prizes/themes/physics/ekspong/.

[23] *See, e.g.*, Avizienis et al., "IEEE Basic concepts and taxonomy of dependable and secure computing," http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1335465.

[24] We intentionally use "risk" in the colloquial sense. *See*, *e.g.,* the Oxford English Dictionary definition, risk, n.: "(Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility." In many frameworks, risk is defined as a level of relative concern, the product of some factors (impact x probability = risk). Those definitions are functional in the context of formal, quantitative or semi-quantitative risk assessment, but should not be confused with our usage here.

1: Comprehensivity

Identify and account for all relevant systems, actors, and risks in the environment.

"Am I covering all of my bases?"

> *If you know your enemies and know yourself, you will not be put at risk even in a hundred battles.*
>
> → **Sun Tzu, *The Art of War*, Chapter III, verse 18**[25]

> *Knowledge isn't power until it is applied.*
>
> → **Dale Carnegie**

You can't defend – or defend against – what you don't know is there. That knowledge drives the deployment of resources through time and space. The Comprehensivity principle sets the scope for information security: the security practitioner must both identify and account for all risks in the environment that may impact their mission. Contemporary information systems are incredibly complex and interconnected, and the environments they must operate in are even more so. This complexity and interconnectedness allows for the compromise of even a single element to potentially compromise the system as a whole, and makes knowing your enemies and knowing yourself both incredibly important and incredibly difficult. To combat these challenges, security practitioners must constantly be on the lookout for potential risks to the mission, and must continuously take steps to ensure that those risks are accounted for, all in striving towards Comprehensivity.

- Cyber attackers' reconnaissance efforts frequently put them at an advantage over defenders, who may be less familiar with the systems and data flows they are tasked with protecting. *In contemporary cybersecurity, knowing thyself means having accurate, up-to-date information asset inventories, network maps, and data flow diagrams.*
- Adequately protecting against the threats in one's environment requires understanding what those threats are, and this information is rarely forthcoming. To truly understand the threats one faces, *practitioners must engage in reconnaissance and intelligence gathering measures for the threats arising from their environment.*
- Inputting backdoors, middlemen, or key-escrow protocols into encryption fundamentally weakens the security that encryption provides, because it opens additional vectors for attackers. *Using end-to-end and full-disk encryption provides the most comprehensive security for encrypted systems.*[26]
- Assuming that users within a given system are authorized to access the content in

[25] *Available at* http://classics.mit.edu/Tzu/artwar.html.

[26] However, encryption is not solely within the domain of Comprehensivity, and our discussion here should not be construed as taking a position on the ongoing public-policy debate over encryption backdoors. As with all of information security, encryption is subject to all of the Principles, including the competing principle of Proportionality, and the benefits offered by end-to-end encryption may ultimately be outweighed by the countervailing needs of the mission, (in this case, collective security through law enforcement). Our purpose here is simply to illuminate the terms on which this debate should be argued, and to inform practitioners of the interests and analyses that must be brought to bear when confronting this issue.

that system provides a venue for attackers to skirt authentication through privilege escalation and other internal system attacks. Rather than relegate authentication solely to the front gate, or rely on assumptions of privilege, *practitioners should use complete mediation to ensure security in authentication throughout their systems*.

- No organization operates in true isolation, and vulnerabilities often arise from manufacturers, suppliers, third party distributors, middlemen, and other outside entities. *Rather than open yourself up to these outside vulnerabilities, practitioners must take steps to secure their supply chain*.

Although Comprehensivity by definition encompases all of information security, it is perhaps best understood in terms of preparedness and prevention. Comprehensivity's most pronounced impact is that risks can be squashed before they ever have a negative impact: where victories are won without a fight, and where losses are prevented rather than mitigated. Many an intelligence victory is never celebrated or even observed, due to foreknowledge and prevention. As such, Comprehensivity is especially important in fields where the practitioner's actions impact the safety, security, and/or well-being of others, and where every failure is critical. The comprehensive information security practitioner rapidly identifies the risks to a campaign, mission, or system, and takes steps to account for those risks before they are given a chance to have a meaningful impact.

However, understanding Comprehensivity at a practical level requires Proportionality and Rigor. Proportionality serves to temper Comprehensivity's overwhelming scope by establishing priorities and allowing for the consideration of competing interests. Unrestrained Comprehensivity can be financially and physically exhausting, and Proportionality emphasizes the necessity of striking the appropriate balance. Rigor, by contrast, plays a key supporting role: by establishing specific procedures and accountability measures, you can help to ensure that the practitioner actually and consistently adheres to Comprehensivity.

Grounding

In the wake of the First World War's brutal trench warfare, the French military undertook a massive project of fortifying the French-German border to ensure French superiority in the case of future German aggression. These fortifications, known collectively as the Maginot Line, provided some of the most thorough and advanced military defenses up to that point in history, spanning the French border with Germany. Yet this strategy left the French border with Belgium unfortified, and particularly the Ardennes Forest strategically unprotected. This weak point in the French defensive line served as the focal point for the German advance in the Battle of France during World War II, allowing the German army to quickly bypass the Maginot Line and rout the French and British armies.

The Maginot Line has gone down in history as one of the great military blunders, where a massive strategic undertaking was rendered completely ineffective due to the proverbial 'weak link.' The French miscalculation of the likelihood of a German invasion through the Ardennes, as well as their failure to anticipate the new German Blitzkrieg-style offensive perfectly encapsulates the importance of Comprehensivity: the strongest defenses are all but worthless if the attacker can simply go around them. This lesson is particularly important in the cyber context, where attackers have proved tremendously successful at finding and exploiting these undefended and under-defended areas to easily bypass existing security controls. Indeed, avoiding a “Cyber Maginot Line” has become something of a rallying cry recently, as it has become increasingly clear both how large the scope of the cyber problem is, and how important comprehensive defenses are in defending against it.[27]

Comprehensivity is therefore primarily concerned with avoiding this “Cyber Maginot Line,” and it does so by ensuring that practitioners fully understand and account for all risks in the environment. Any security assumption or oversight can serve as the proverbial Ardennes, so practitioners must learn to think comprehensively about the mission they are tasked with defending, to proactively identify emerging or future threats, and to develop strategies that fully account for the risks facing their systems.

Strategies & Challenges

Implementing Comprehensivity can be thought of broadly in three steps: (1) The practitioner must comprehensively *identify* the risks facing a particular system; (2) The practitioner must develop security controls or security strategies that adequately *account* for each of the identified risks; and (3) The practitioner must continuously re-engage in steps (1) and (2), identifying new risks and creating new strategies as necessary. Considering the cyclical nature of Comprehensivity, it requires diligence,[28] vigilance,[29] and proactivity.[30]

The first step, *identify*, requires the security practitioner to have a robust understanding of the system they are tasked with defending, and the environment that system will operate in. This begins with standard business processes like *inventorying* and *network mapping*, but also requires environmental assessments through *reconnaissance* and *information sharing*. And when operating at an organizational level, the information gathered must be disseminated through *awareness-raising* measures and *training*.[31] Due to the importance of process in executing the *identify* step, Comprehensivity is closely tied to Rigor.

[27] “Cybersecurity's Maginot Line: A Real-world Assessment of the Defense-in-Depth Model,” FireEye, Inc., https://www2.fireeye.com/real-world-assessment.html.

[28] *See, e.g.*, Annegret Bandiek, “Due Diligence in Cyberspace: Guidelines for International and European Cyber Policy and Cybersecurity Policy,” SWP, (May 2016), https://www.swp-berlin.org/fileadmin/contents/products/research_papers/2016RP07_bdk.pdf.

[29] *See, e.g.,* John Kellogg, “CyberSecurity begins with diligence, requires continued vigilance,” US Army, (Oct. 26, 2015), https://www.army.mil/article/157693/CyberSecurity_begins_with_diligence__requires_continued_vigilance.

[30] *See, e.g.*, Larry Karisny, “Cybersecurity: Taking a Proactive Approach is key,” GovTech, (March 4, 2015), http://www.govtech.com/dc/articles/Cybersecurity-Taking-a-Proactive-Approach-is-Key.html.

[31] Although awareness-raising measures often operate on systems with much greater organizational complexity, (e.g. entire businesses), they reflect the same underlying principle: the security of a given system is fundamentally undermined if it fails to identify and account for all the potential risks arising from its environment. Current examples like supply chain security and BYOD are attempts to identify an existing blind-spot in organizational security strategies, and encourage security practitioners to adequately account for them.

> *However, the identify step should not be oversimplified into a rote itemization of threats and assets. The identify step represents the need for the practitioner to "know thyself" and "know thy enemy," and as such entails a detailed and nuanced perspective on these issues. As such, the practitioner should understand and utilize methods that allow for the visualization of not just system elements, but the architecture of how those elements interact, as well as to understand and utilize methods for modeling threats in a manner that clearly conveys the relative risks they present.*[32]

The second step, *account,* requires that the risks identified are appropriately protected against. It is important to note that the word "account" is not synonymous with "defend": every risk identified does not necessarily require a unique security control. Rather, the risks must be accounted for. This may be accomplished through security strategies that manage multiple risks, or through the determination that certain risks are acceptable to the organization's overall mission. Due to the close relationship between accounting for risks and understanding the relative importance of those risks in the context of the broader mission, the *account* step is closely tied to Proportionality.

> *Nevertheless, adequately accounting for the risks facing a system often requires complete protection across that system, as seen with best practices such as: complete mediation*[33] *with authentication; end-to-end encryption*[34] *for data transit and communications; and lifecycle planning for product deployment. Non-comprehensive application of security in any of these areas leaves a security gap which may undermine the security of the larger system.*[35] *So while Comprehensivity provides room for security strategies to take strategic risks, in many cases Comprehensivity ultimately requires complete protections.*

The final step, *diligence*, teaches that Comprehensivity applies not just to space, but also to time. New threats emerge daily, old systems and strategies become obsolete, organizations inherit or purchase systems with pre-existing vulnerabilities, and opportunities may arise to further enhance security. Therefore, proper implementation of the Comprehensivity principle requires the practitioner to consider risks that may arise in the future, or may stem from far in the past. On the micro-scale, this may include *continuous monitoring*[36] and

[32] For an in-depth discussion of threat modeling, including the creation of threat trees to prioritize and manage risks, *see, e.g.,* Adam Shostack, Threat Modeling: Designing for Security, Wiley, (February 2014), ISBN: 978-1-118-80999-0.

[33] Complete Mediation is defined as: "Every access to every object must be checked for authority." *See, e.g.,* Saltzer and Schroeder.

[34] End-to-end Encryption is defined as: "Encryption of information at its origin and decryption at its intended destination without intermediate decryption." National Information Assurance Glossary, CNSSI 4009, https://cryptosmith.files.wordpress.com/2015/08/glossary-2015-cnss.pdf. *See also*, Andy Greenberg, "Hacker Lexicon: What is End-to-End Encryption? Wired," (Nov. 25, 2014), https://www.wired.com/2014/11/hacker-lexicon-end-to-end-encryption/.

[35] *See, e.g.,* Abelson et al., "Keys Under Doormats: Mandating insecurity by requiring government access to all data and communications," MIT, (July 6, 2015), *available at* http://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf?sequence=8 (discussing insecurity imposed by mandatory backdoors in lieu of end-to-end encryption).

[36] Continuous Monitoring is defined as: "Maintaining ongoing awareness of information security, vulnerabilities, and threats to support organizational risk management decisions." Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations, NIST 800-137, http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-137.pdf.

situational awareness[37] measures. Yet practitioners must also implement security strategies that see the long game, and do not succumb to the metaphorical game of vulnerability "Whac-A-Mole." These macro-scale strategies include lifecycle planning, periodic review and maintenance, supply chain and trust evaluation, and investments in infrastructure to ensure that old systems and strategies remain effective and that new systems incorporate cybersecurity as a fundamental consideration.[38]

A major challenge the practitioner is likely to face when implementing Comprehensivity is in identifying and accounting for risks arising from dependencies and other actors. Vulnerabilities arising from dependencies are difficult to identify because the underlying systems are often hidden from the practitioner. Likewise, accounting for these vulnerabilities is often even more difficult, as the practitioner cannot exercise effective control over outside entities. Even when working with relatively open sources, it is easy to overlook old code, commonly used libraries, and unassuming middlemen as potential points of vulnerability. While greater control over these dependencies can be exercised through contract, or can be removed entirely by internalizing high-risk activities, most will require periodic reassessment, and all will require proactivity when major vulnerabilities are discovered.[39]

[37] Situational Awareness is defined as: "Knowledge and understanding of the current situation which promotes timely, relevant and accurate assessment of friendly, enemy and other operations within the battle space in order to facilitate decision-making." U.S. Army Field Manual - Operational Terms and Graphics, FM 1-02, MCRP 5-12A, http://www1.udel.edu/armyrotc/current_cadets/cadet_resources/manuals_regulations_files/FM%201-02%20-%20Operational%20Terms%20&%20Graphics.pdf.

[38] For a high level discussion of security by design, *see, e.g.*, Security by Design Principles, OWASP, https://www.owasp.org/index.php/Security_by_Design_Principles.

[39] *See, e.g.*, Imagemagick Vulnerability, US-CERT, https://www.us-cert.gov/ncas/current-activity/2016/05/04/ImageMagick-Vulnerability.

2: Opportunity

Take advantage of the actor relationships, material resources, and strategic opportunities available in the environment.

"Am I taking advantage of my environment?"

> *Unhappy is the fate of one who tries to win his battles and succeed in his attacks without cultivating the spirit of enterprise; for the result is waste of time and general stagnation. Hence the saying: The enlightened ruler lays his plans well ahead; the good general cultivates his resources.*
>
> → **Sun Tzu, *The Art of War*, Chapter XII, verses 15-16**[40]

> *There is no security on this earth; there is only opportunity.*
>
> → **Douglas MacArthur**

The environment is more than a source of threats, it is a source of opportunities. Too often security practitioners thwart themselves by wasting resources, missing out on important information, or otherwise suffering the losses of operating in an information silo. Although the costs of ignoring the Opportunity principle are often more visible than the gains to be had in applying it, practitioners should look beyond their borders and seek to harness the opportunities available in their environment:

- Maintaining a custom authentication solution is costly and risky, compared to contributing to the development of a standardized solution backed by many organizations, with protocols and source code vetted by many eyes. *Using and supporting common tools reduces the burden of doing cybersecurity well.*
- Exploit kits in the black hat world spread like wildfire, often seeing mass adoption within hours of publication. White hats, however, are notoriously slow to share information, and often only in formalized settings out of fear of liability or criticism. This disparity in response times gives attackers a substantial edge, with the good guys responding on a timescale of days, compared to the bad guys' hours. *Effective information sharing decreases response time, and makes defenses more effective.*[41]
- Honeypots[42] and canaries[43] can be extremely useful in high-security environments for detecting adversarial actions before they have reached critical systems or data. *Deceiving an attacker with false targets is a powerful tactic for more mature*

[40] *See* Sun Tzu, *supra*.

[41] *See* VADM A. K. Cebroski and J. J. Garstka, *Network-Center Warfare: Its Origin and Future*, Proceedings, (January 1998), available: http://www.kinection.com/ncoic/ncw_origin_future.pdf, highlighting the "shift from viewing actors as independent to viewing them as part of a continuously adapting ecosystem."

[42] A honeypot is a computer mechanism designed to detect, monitor, deflect, or counteract an attempt by an attacker to gain unauthorized access to a computer system. For a more complete discussion of honeypots, *see, e.g.,* N. C. Rowe and H. C. Goh, "Thwarting Cyber-Attack Reconnaissance with Inconsistency and Deception," *Information Assurance and Security Workshop, 2007. IAW '07. IEEE SMC*, West Point, NY, (2007), pp. 151-158. doi: 10.1109/IAW.2007.381927.

[43] A canary is a slang term for a computer intrusion detection system that can identify malicious behavior.

cybersecurity programs.

- Penetration testing – simulated break-ins by a trusted analyst – can reveal holes in defenses before an adversary finds them. Combined with live security exercises, these techniques can enhance process, morale, and skills among responders and other stakeholders. *Taking advantage of offensive tools and techniques provides powerful leverage for identifying and communicating risks and vulnerabilities before they become disasters.*

Unlike Comprehensivity, which is focused on scope, knowing what you have, and planning for the long game, Opportunity is about proactively seeking and cultivating resources and advantages, including: tools, collaborations, openings for deception, information sharing, and trust. Opportunity tells the practitioner to look beyond the bounds of their specific system, to view the environment as a tool to be harnessed, and to understand that effective security cannot be achieved through total isolation. Attackers already harness the power of connectivity: security practitioners should do the same.

To this end, the impact from failing to implement Opportunity is just that: missed opportunities. Unfortunately, missed opportunities are inherently difficult to quantify, and any attempts to do so will be skewed by status quo bias, which negatively predisposes most practitioners towards positive risk,[44] and hindsight bias, where outcomes seem predictable once they've already occurred. Rather ironically, the *potential* vulnerabilities presented by new opportunities are often given far greater weight than the *known* vulnerabilities that exist already in the practitioner's system. To combat this, practitioners must learn to view opportunities in terms of potential gain, and not allow the potential risks these opportunities present to be given disproportionate weight.[45] And as with all of the principles, Proportionality provides the calculus for determining whether a given opportunity is "worth it"; but Opportunity is unique in that it explicitly requires practitioners to apply this calculus toward positive risk taking.

Grounding

The impact that stealth technology had on the outcome of the Cold War is widely acknowledged. What few realize is that America got the idea from a Russian scientist, publishing in a Russian journal. In *Skunk Works*, Ben Rich describes an exceptional young mathematician and radar specialist on his staff, Denys Overholser, bringing him a dense technical paper, "Method of Edge Waves in the Physical Theory of Diffraction," by Pyotr Ufimtsev, chief scientist at the Moscow Institute of Radio Engineering. In the obtuse language of pure research, (as opposed to engineering), Ufimtsev had built on earlier

[44] Status quo bias refers to a bias in human decision-making that uses the current state of affairs as a reference point to which alternatives are viewed disproportionately negatively. *See, e.g.,* William Samuelson & Richard Zeckhauser, "Status Quo Bias in decision-making," Journal of Risk and Uncertainty 1: 7-59, (1988), *available at* https://www.hks.harvard.edu/fs/rzeckhau/SQBDM.pdf. For a discussion of status quo bias in information security, *see, e.g.,* Shari L. Pfleeger & Deanna D. Caputo, "Leveraging Behavioral Science to Mitigate Cybersecurity Risk," MITRE Corp., *available at* https://ai2-s2-pdfs.s3.amazonaws.com/e755/aa8baf01ef655ef7b1472ceba505b7c45b91.pdf.

[45] *See, e.g.*, Katherine Noyes, "10 Reasons Open Source is Good for Business," PCWorld, (Nov. 5, 2010), http://www.pcworld.com/article/209891/10_reasons_open_source_is_good_for_business.html.

works to describe an accurate method for calculating radar reflection and refraction off of polygonal surfaces.

In other words: Ufimtsev had told anyone nerdy enough to delve into his paper (probably long forgotten in a filing cabinet somewhere in Russia) how to make something disappear off of radar. This is the Opportunity principle at work: the Russians failed to create a pipeline for scientific research to make it into engineers' hands in a digestible form, so they missed out on stealth. America was no better at this in general, but one man made a difference by connecting himself outside the intellectual silos common to his workplace, and finding a treasure.

Information security is enhanced by the Opportunity principle in similar ways. By using tools that are well-resourced and vetted by a wide community, such as open source cryptography libraries, compilers, and virtualization tools, security practitioners can benefit from the added scrutiny of a larger community.[46] By sharing information with other practitioners and organizations,[47,48] we evolve the speed and quality of our responses to a changing environment.[49,50] This connectedness can benefit organizations in other ways as well, by having a ready hiring pipeline when personnel turnover happens, by having outsiders to bounce ideas off of for fresh perspectives on strategy or a new analysis on a technical control, and more.

Strategies & Challenges

The Opportunity principle in many ways mirrors Comprehensivity: practitioners must identify, evaluate, and act upon[51] the opportunities in their environment. Broadly speaking, opportunities arise in three categories: (1) actor relationships; (2) material resources; and (3) strategic opportunities. Put another way, opportunities arise in relation to the three primary definitional areas: actors, systems, and strategies. While instantiations of these opportunities will certainly vary greatly, even within a given category, it is important to understand that opportunities can impact all aspects of an organization's security, and that organizations should seek to harness these wide-ranging benefits.

The first category, *actor relationships*, is concerned with harnessing the aligning or potentially aligning missions of other actors within your environment. Practitioners shouldn't "go-it alone," and the combined resources of a coalition of actors will always outshine any individual. At a practical level, this may simply involve enlisting the services of outside security firms, such as for independent auditing and penetration testing in larger

[46] *See, e.g.*, Linus' Law ("With multiple eyes, all bugs are shallow.")
[47] *See, e.g.,* DoD Cyber Strategy 2015, Strategic goal #5 "Build and maintain robust international alliances and partnerships to deter shared threats and increase international security and stability." *See also*, "joint operations" in military strategy. Army Field Manual, Appendix A, https://fas.org/irp/doddir/army/fm3-0.pdf; Canadian Principles of War - "Cooperation" http://publications.gc.ca/collections/collection_2010/forces/D2-252-2009-eng.pdf.
[48] *See* DoD Cyber Strategy, "information sharing and interagency coordination."
[49] *See, e.g.,* McGraw, *supra* ("Borrow others' good ideas")
[50] *See, e.g.,* Cybersecurity National Action Plan, (Feb. 9, 2016), *available at* https://www.whitehouse.gov/the-press-office/2016/02/09/fact-sheet-cybersecurity-national-action-plan.
[51] Or not act upon, as the case may be.

organizations, or completely outsourcing elements of security to third-party *Security as a Service* vendors for smaller ones.[52] More broadly, this may also include participating in information sharing organizations between public and private actors, as well as developing deeper organizational ties between like-minded organizations. Even nation-states benefit from strategic partnerships with state and non-state actors, ranging from participating in information sharing organizations[53] to conducting wargames and tabletop exercises.[54]

The second category, *material resources*, represents the vast array of information security resources that are either publically available or widely adopted across the information security landscape. Practitioners need not reinvent the wheel, particularly when the existing wheel is thoroughly tested and widely adopted.[55] Open source tools and standards are secure and cheap, and reliance on these widely adopted tools and standards eases learnability, interoperability, and testability. Whereas organizations of any size can benefit from sending personnel to security conferences[56] and ensuring they have access to a variety of security publications. While some security requirements will be too specialized or sensitive for an open source solution, and novel problems may not be well known in the broader community, all practitioners can utilize these public resources to some degree, and will benefit from designing tools with broader standards and practices in mind.

The final category, *strategic opportunities*, recommends that security practitioners take a more active role in the defense of their system. Rather than view security purely as a passive pursuit, practitioners should embrace proactive, innovative, and aggressive solutions, and thereby manipulate their environment to further the needs of their mission. These strategic opportunities are often more offensive and adversarial, and include activities such as deterrence, active defense, and deception.[57] However, strategic opportunities are not purely adversarial, and may involve appeals to shared interests or diplomacy, particularly at the state-actor level.[58] In all cases, the important concept is that

[52] *See, e.g.,* Kevin M. Henry. *Penetration Testing: Protecting Networks and Systems*. IT Governance Ltd. ISBN 978-1-849-28371-7. ("Penetration testing is the simulation of an attack on a system, network, piece of equipment or other facility, with the objective of proving how vulnerable that system or 'target' would be to a real attack.")

[53] *See, e.g.,* Presidential Decision Directive 63, *available at* http://fas.org/irp/offdocs/pdd/pdd-63.htm; Executive Order 13691, *available at* https://www.whitehouse.gov/the-press-office/2015/02/13/executive-order-promoting-private-sector-cybersecurity-information-shari; For an example ISAC, *see, e.g.,* FS-ISAC https://www.fsisac.com/about.

[54] *See, e.g.,* Heather Stewart, "UK and US to simulate cyber-attack on nuclear plants to test resilience," The Guardian, (Mar. 20, 2016), https://www.theguardian.com/uk-news/2016/mar/31/uk-us-simulate-cyber-attack-nuclear-plants-test-resilience.

[55] *See, e.g.,* "TLS and SSL: A Beginner's Guide," SANS, https://www.sans.org/reading-room/whitepapers/protocols/ssl-tls-beginners-guide-1029.

[56] Common examples of security conferences are RSA and Blackhat, both of which attract a wide range of security professionals annually.

[57] *See, e.g.,* Josh Johnson, "Implementing Active Defense Systems on Private Networks," SANS, (2013), *available at* https://www.sans.org/reading-room/whitepapers/detection/implementing-active-defense-systems-private-networks-34312.

[58] *See, e.g.,* Joseph Menn and Jim Finkle, "Chinese economic cyber-espionage plummets in US - experts," Reuters, (June 25, 2016), http://www.reuters.com/article/us-cyber-spying-china-idUSKCN0Z700D (discussing the decline in

the practitioner is seeking to actively harness characteristics of the environment for their own strategic advantage.

> *It is important to note, however, that Opportunity's greater impact on the surrounding environment means it is also the most likely principle to come in conflict with the rules, laws, and norms governing those environments.[59] For instance, although deterrence is certainly vital to enhancing security, it also presents the threat of escalation, and reckless application of Opportunity can certainly do more harm than good. As such, Opportunity requires the most careful scrutiny of all of the Principles, and must always be cautioned against engaging in behavior that is unlawful or disproportionately damaging relative to the threats faced.*

In addition to these legal difficulties, opportunities also present many practical difficulties. After all, opportunities frequently operate on new or untested ground, and understanding the relative benefits of a given opportunity can be both difficult and controversial. New technologies are often difficult to evaluate due to scarce evidence,[60] and even long standing security debates present challenges when the relative merits are hotly contested. For example, the perennial "openness vs. secrecy" debate has been litigated *ad nauseum*, yet there is little in the way of consensus, as different parties ultimately place different values on the relative benefits and burdens secrecy and openness present.[61] While the Opportunity principle does not directly take a side on these issues, it does require the practitioner to be aware of the benefits and burdens that new opportunities may provide, and to be able to draw reasonable conclusions based upon the evidence available and the needs of the mission.[62]

Chinse cyber-espionage in response to US-China diplomacy).

[59] For instance, for a discussion of the legality of active defense, *see* "Into the Grey Zone: The Private Sector and Active Defense against Cyber Threats," Center for Cyber and Homeland Security, GWU, (Oct. 31, 2016), *available at* https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/CCHS-ActiveDefenseReportFINAL.pdf.

[60] The issue of decision-making based on limited evidence is discussed in greater detail with regard to Rigor.

[61] *See, e.g.,* Bruce Schneier, "The Non-Security of Secrecy," (Oct. 2004), *available at* https://www.schneier.com/essays/archives/2004/10/the_non-security_of.html; *contast with*: A. Schmidt, "Secrecy versus openness: Internet security and the limits of open source and peer production," (Nov. 3, 2014), http://repository.tudelft.nl/islandora/object/uuid:ecf237ed-7131-4455-917f-11e55e03df0d/?collection=research.

[62] For a more in-depth discussion on this multi-factor calculus, *see* Proportionality.

3: Rigor

Specify and enforce the expected states, behaviors, and processes governing the relevant systems and actors.

"What is correct behavior, and how am I ensuring it?"

> *The last rule was to make enumerations so complete, and reviews so comprehensive, that I should be certain of omitting nothing.*
>
> **→ René Descartes, *Discourse on the Method*[63]**

> *"Ostinato Rigore" (Constant Rigor)*
>
> **→ Leonardo da Vinci, Personal Motto**

Information security doesn't work when approached in a haphazard or disorganized manner: it must be *rigorous*.[64] Assumptions must be enumerated; procedures must be articulated; requirements must be explicit; and accountability criteria must be enforced. Rigor, therefore, is primarily concerned with process, and specifically those processes designed to establish and ensure the success of the mission. Rigor seeks to make outcomes predictable and failures diagnosable after the fact. While no implementation will ever be perfect, adherence to the Rigor principle ensures that best practices will be consistently applied, that security incidents will be efficiently responded to, and that failures will be used to improve security outcomes in the future.

- Information systems require the same mechanisms for internal governance that are seen in politics, international relations, and corporations. Rather than assume cyber-governance issues will solve themselves, *practitioners should put in place a robust system of rules, authorities, responsibilities, and consequences for their information systems.*
- Software projects that begin with unclear goals and ambiguous tasks often create buggy code, overlapping functions, and vulnerabilities. Rather than approach projects in a haphazard manner, *practitioners should design their systems based on clear requirements, identifying the expected inputs and outputs for each element.*
- Even the best designed security systems can have flaws, oversights, and internal failings. Rather than assume that your systems will work as intended, *practitioners must ensure that their systems are audited, that their policies are adhered to, and that violations face enforcement.*[65]

[63] This rule represents the fourth and final step for seeking knowledge articulated by Descartes in his Discourse on the Method. The purpose of the four rules was to create a process whereby he could ascertain truth, to the greatest degree possible, notwithstanding physical limitations.

[64] Our intent in using the word "rigor" is not to invoke the cold, lifeless meaning found in medicine, as in "rigor mortis," but rather the definition used in scientific literature more generally. Science must be thorough, exacting, and precise, but not harsh or unfeeling. Security practitioners too must adhere to this kind of rigor. For instance, OED defines "rigor" as "the quality of being extremely thorough and careful." Oxford English Dictionary, *available at* https://en.oxforddictionaries.com/definition/rigour.

[65] For an in-depth discussion of Audits in the information security context, *see, e.g.*, ISACA IT Assurance Framework (ITAF), *available at*

- Decision-making processes are only as good as the information that informs them, and in today's information systems, that information must be acquired quickly, and evaluated efficiently. To ensure that these important decisions have up-to-date and accurate information, *practitioners should put in place mechanisms to monitor security sensitive activities and provide a constant flow of security critical data.*

Considering Rigor's role as the "process and governance principle," this principle is particularly important to understand in combination with Comprehensivity, Opportunity, and Proportionality. Comprehensivity is important for forming the information-base which informs rigorous decision-making: you cannot employ rigorous standards to your software dependencies if you don't even know whose code you're running. Opportunity and Proportionality are important in combination with Rigor for finding the appropriate balance between prescriptive and open-ended requirements. The Opportunity principle thrives on lightweight and thrifty implementations of Rigor, as the practitioner can capitalize on new opportunities without getting bogged down in paperwork and procedure. But if Rigor becomes too lightweight, it risks losing its value altogether, where guidelines are written but never followed, or enforcement is so toothless as to never deter bad behavior. As such, Proportionality is important for striking the appropriate balance between overly strict and overly lax implementations of Rigor.

The impact of the Rigor principle is two-fold: (1) Rigor better ensures positive security outcomes, as actors have clearly defined roles, responsibilities, and expectations; and (2) Rigor allows for the more effective identification of and response to security failings. Responses to some specific failings will already be established, and diagnosing and rectifying the underlying problems will be facilitated by a preexisting structure. Thought of another way, the failure to practice Rigor can lead to procedural oversights, ambiguity with regard to responsibilities, ineffective evaluations, and clumsy incident response.

Grounding

To elucidate the importance of process in ensuring positive outcomes, it is helpful to look to an outside field: surgery. In the book *The Checklist Manifesto,* Harvard professor and surgeon Atul Gawande paints a troubling picture of the state of surgical assurance. Surgeons, as Gawande describes them, tend to view themselves as rockstars in the operating room, expecting absolute deference to their skills and discretion to ensure patient safety. And to some degree this makes sense: surgery is complicated, and a lot of things can go wrong; surely you shouldn't try to predetermine the surgeon's action when there are so many contingencies? Yet Gawande goes on to prove just how wrong this assumption can be. Surgeons, it turns out, are just as fallible the rest of us, and relying purely on their instincts can lead to common, yet impactful, mistakes. To this end, Gawande shows how substantial gains in patient outcomes can be obtained through the implementation of a simple baseline of procedurality - a checklist.

http://www.isaca.org/Knowledge-Center/Research/Documents/ITAF-3rd-Edition_fmk_Eng_1014.pdf.

> The results Gawande describes from adherence to his checklist are staggering: 36% reduction in major complications; 47% reduction in patient deaths; 50% reduction in postoperative infections.[66] All from adherence to a 19-step checklist. This list was not designed to encompass the vast array of surgical complications that may arise. Rather, it was designed to identify the most impactful and frequently skipped steps to ensure patient safety. The items on the list – appropriate timing of antibiotics; proper hygiene and cleanliness; building out communication and teamwork among the surgical team – were not complex or unknown: rather, they were among the most basic things doctors are taught. Yet by taking steps to *ensure* that these most basic procedures were followed, hospitals were able to substantially improve patient outcomes, and by extension, reduce costs.
>
> The lessons Gawande describe are paramount to understanding Rigor. The role of the checklist is not to do the surgeon's job for them: the checklist is a tool to aid the surgeon in doing their job more effectively. Indeed, across a range of fields, from logistics to aviation to structural engineering, one of the most important traits Gawande identifies for an effective checklist is to make room for experts to do their job, and perhaps more importantly, communicate with one another.[67] By understanding this fundamental balance – between prescribing action and imbuing flexibility – security practitioners will be able to create the processes to ensure appropriate action, while allowing for the freedom to actually carry out those actions. For instance, when crafting incident response "playbooks," those playbooks need to be clear about who has authority, what each actors' responsibilities are, and what steps need to be taken in response to specific incidents. But playbooks can never predict every future scenario, and thus must ensure that people with expertise and experience are empowered to make critical decisions when a best path forward is not set out in advance.

Strategies & Challenges

Implementation of Rigor reflects an ongoing process: the procedures Rigor puts in place are not static: they are routinely updated in response to environmental changes and internal evaluations. While the specific structure may vary from common tools – like PDCA cycles[68] and OODA loops[69] – to more specialized procedures developed for specific tasks,

[66] Gawande's study was conducted in hospitals ranging from advanced Western hospitals to some of the poorest in the world. In all instances, adherence to the checklist provided substantial benefits, suggesting the problem identified was systemic, not local to any particular hospital or medical culture.

[67] Gawande also emphasizes the importance of updating the checklist based on evidence and experience, testing each iteration of the checklist to assess the impact it has on outcomes, and disseminating the information gathered to a wide variety of stakeholders and participants.

[68] A PDCA loop is an iterative process for task management consisting of four steps: Plan-Do-Check-Act. For a more complete discussion of the PDCA cycle and their usage in task management, *see, e.g*., "Plan-Do-Check Act Cycle," American Society for Quality, http://asq.org/learn-about-quality/project-planning-tools/overview/pdca-cycle.html.

[69] OODA loops are decision cycles consisting of the four steps: Observe, Orient, Decide, Act. For a more complete discussion of OODA loops, see, e.g., John Boyd, "The Essence of Winning and Losing,"

they should all reflect the ongoing, self-assessing, and self-updating nature of Rigor. Rigor exists largely in two phases: (1) *specification,* and (2) *enforcement*. Specification can be thought of as the "write it down" phase, where the ownership, requirements, responsibilities, and assumptions of specific systems are enumerated. Enforcement, by contrast, is the "go do it" phase, where the specifications are verified, implemented, evaluated, and updated. While these two phases can be further subdivided to suit the practitioner's needs, this two-phase structure provides a valuable baseline for practitioners to begin implementing Rigor.

The first step, *specification*, is where the practitioner enumerates specific substantive requirements that are required to adequately govern a given system. By specifying, the practitioner leaves little to assumption, and ensures that all parties going forward understand the essential governing aspects for each system they are tasked with defending. Although highly contingent on the needs of the particular system, specification will typically benefit from a statement of the system's goals or mission; statements of ownership and authority, both within the system and with regard to other systems; statements of roles and responsibilities for the various actors and elements that interact with that system; and a statement of how the system will evaluate and improve upon itself.

> *Looked at from a broader perspective, specification is about putting in place the necessary pieces for a system to operate in a rigorous manner. Specification is where the practitioner establishes restrictions and processes necessary to ensure the success of their mission. This ranges from clarifying potential ambiguities and facilitating communication, to enabling visibility into the internal working of the system, to creating processes for self-improvement when faults are identified. While specification will mean different things to different systems, at its core, it's about asking "what do I need to ensure the success of my mission?", and putting those things in place.*

The second step, *enforcement*, focuses on verifying, implementing, evaluating, and updating the substantive requirements created during specification. At its most basic, enforcement is where departures from the requirements of specification are set right. But enforcement goes further, and requires the practitioner to evaluate whether or not the controls and procedures specified are accomplishing what they set out to do, and to make alterations where necessary to ensure better security. However, haphazard or sloppy enforcement should not be considered sufficient: enforcement too must be rigorous. This requires that the enforcements be made on the basis of evidence, utilizing meaningful security metrics whenever possible, and that the enforcement process is clear, transparent, and impartial.[70]

The emphasis on security metrics and evidence-based decision-making, although critical, will also likely serve as one of the greatest challenges in implementing the Rigor principle in

https://web.archive.org/web/20110324054054/http://www.danford.net/boyd/essence.htm.

[70] Fairness in decision making is an incredibly complex topic, and cannot be fully addressed within the scope of this paper. Prominent jurisprudential examples include the Due Process Clause of the U.S. Constitutional, the concept of "fundamental justice" in Canadian and New Zealand law, and "procedural fairness" in Australian law. For an influential discussion of the requirements of fairness in decision making, *see* Henry J. Friendly, "Some Kind of Hearing," 123 U. Pa. L. Rev. 1267 (1975).

practice. Robust metrics for evaluating security, although frequently discussed,[71] have proven difficult to identify, and those that have been identified often have dubious value to improving security.[72] Yet practitioners should not let this lack of gold-standard evidence deter innovation or investment in emerging technologies, as many valuable controls may have little or no empirical evidence supporting their use.[73] Rather than live or die by a single study, practitioners should employ a Bayesian approach[74] of analytical thinking, where new evidence informs, but does not invariably determine the practitioner's course of action. Rather, new evidence should be given its appropriate evidentiary weight, and combined with well-founded prior beliefs, established best practices, and expert advice to help identify practices that will best ensure security outcomes.[75]

Finally, an important point of clarification for Rigor is that "evaluation and accountability" should not be reduced to mere scapegoating. Although major security failings often necessitate stern accountability measures, potentially including firings, this is rarely a sufficient response. The importance of evaluation and accountability is in ensuring that security failings and oversights are properly remedied. Although this may simply amount to one person failing to do their job, the more likely cause will be systemic shortcomings that had not previously been identified or acted upon. Therefore, evaluation and accountability are most important for their role in reframing existing policies and procedures to rectify the shortcomings the security failing identified. Since a single policy will never be perfect, the Rigor principle is designed to reflect the ongoing nature of security policy development.

[71] *See, e.g.*, Black et al., Cyber Security Metrics and Measures, NIST, (Mar. 2, 2009), *available at* https://www.nist.gov/publications/cyber-security-metrics-and-measures.

[72] Jai Vijayan, The Four Big Problems with Security Metrics, DarkReading, (Jan. 11, 2016), http://www.darkreading.com/risk/the-four-big-problems-with-security-metrics/d/d-id/1323849.

[73] For an example of more evidence-based cybersecurity recommendations, *see* "Top 4 Mitigation Strategies to protect your ICT systems," Australian Government Cyber Security Operations Center, (October 2012), *available at* http://www.asd.gov.au/publications/protect/Top_4_Mitigations.pdf (finding that: (1) application whitelisting, (2) patching, (3) restricting admin privileges, and (4) defense in depth, were sufficient to stop 85% of cyber-intrusions.)

[74] Bayesian inference refers to a method of statistical analysis that incorporates new evidence with prior beliefs to update the overall probability of a given outcome. Bayesian inference proves particularly important in areas with limited or emerging evidence, and has proved invaluable across numerous fields, including science, engineering, philosophy, medicine, and law. *See, e.g.,* Fienberg & Schervish, The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking, 66 B.U. L. Rev. 771 (1986).

[75] From a legal evidentiary perspective, this is loosely equivalent to the concept of the "weight of the evidence."

4: Minimization

Minimize the size, quantity, and complexity of what is to be protected, and limit externally facing points of attack.

"Can this be a smaller target?"

> *The trodden worm curls up. This testifies to its caution. It thus reduces its chances of being trodden upon again. In the language of morality: Humility.*
>
> → **Friedrich Nietzsche, *Twilight of the Idols***

> *In order for three people to keep a secret, two must be dead.*
>
> → **Benjamin Franklin**

If the world of information security can ever take lessons from the animal world, it will most likely be found in the Minimization principle. From the armadillo rolling into a ball, to the turtle hiding in its shell, nature is replete with animals that defend against predators by becoming smaller, more compact targets. If there is less to attack, the target is less appealing to attackers; if the target's assets are constrained, it is less valuable to attackers; and if the target's design is simple, it will present fewer flaws for attackers to exploit.[76] Smaller and simpler targets cost less to defend, are easier to understand, and will have fewer seams and chinks. Although the trend of information systems is toward growth and complexity, maintaining those systems in a minimized state will ensure that they are easier to secure, and more adaptable to changes in their environment.

- Collecting user Social Security Numbers (SSNs) or other security-sensitive information when not strictly necessary (or keeping it around after it is no longer necessary) makes the user database a valuable target for identity thieves, recruiting a class of attackers that may have ignored the system otherwise. *Holding less information makes you a smaller, less valuable target.*
- Removing an unnecessary interface is a powerful security tactic. Every interface is a point of ingress that an attacker can try to exploit, and that the defender must protect and monitor. *Reduce interfaces to reduce attack surface and focus resources.*
- Software developers often leave unused code in their code base, trusting to its unreachability, tendency to be compiled out, or tendency not to be run. Yet this code may still be the source of vulnerabilities, particularly as other code evolves without it. *Minimizing a system's "moving parts" minimizes both the number of ways the system can fail and the number of ways the system can be attacked.*
- A frequent source of vulnerabilities is that the scale and complexity of your infrastructure outruns your security team's size and resources. When you can no longer keep up with the rate of growth, you will no longer be secure*. Keep the scale and complexity of what must be secured in line with available resources to secure it.*
- Giving human and machine actors within an organization no more access than they absolutely need reduces the impact of any one person's credentials or devices being compromised. *Practice least privilege so that no person's or system's credentials are more valuable than they have to be.*

[76] *See, e.g.*, John Maeda, The Laws of Simplicity.

- User control over what code can be run on a given machine opens up an enormous range of attacks through malicious email attachments, suspicious websites, and simply unknown vulnerabilities. Rather than only blocking code that is known to be bad, (referred to as “default permit” or “blacklisting”), *practitioners should implement application whitelisting, where only approved code can be run on a machine.*[77]

Minimization is the principle that reflects the old adage, “less is more.” Often in the quest for more capabilities, including more *security* capabilities, it can become easy to lose sight of the complexity being introduced. Indeed, Peiter “Mudge” Zatko, in a talk at CanSecWest,[78] pointed out that 28 percent of software vulnerabilities released by dedicated security firms were for flaws in *defensive security products*. Industry’s rush to add more capabilities was in fact introducing more risk, which could quite possibly outweigh any benefit the added capabilities provided. Minimization is about avoiding this needless complexity, and in doing so eliminating unnecessary vulnerabilities, minimizing your value as a target, and increasing the speed at which your systems can adapt to changes internally and to changes in the environment.

Although one of the most intuitive principles, Minimization rarely enjoys a straightforward implementation. While simple in isolation, Minimization has complex relationships with the other Principles, and these tensions require careful consideration to craft systems and strategies that account for all of the Principles. For instance, Comprehensivity, Opportunity, and Rigor all tend to benefit from maintaining systems in smaller, simpler states, as this limits their scope, eases adaptation, and streamlines processes, respectively. Yet Minimization frequently comes into tension with Compartmentation and Fault Tolerance, as both push toward a greater degree of system complexity. Proportionality often requires practitioners to sacrifice some Minimization to accommodate the needs of the mission to grow and expand. Therefore, it is the role of the practitioner to balance these competing concerns, applying Minimization in a manner that accomplishes its goals while advancing the needs of the mission.

Grounding

In 2010, a group of scientists discovered a massive secret hiding in the otherwise unassuming Japanese white flower *Paris japonica*. Hidden in plain sight, this small flower possessed the largest genome ever discovered, an astonishing 150 billion base pairs (nearly 50 times as long as the human genome). Unfortunately, when it comes to genomes, bigger does not mean better. To quote researcher Ilia Leitch, ‟research has demonstrated that those [organisms] with large genomes are at greater risk of extinction, are less

[77] “Whitelisting” refers to an information system architecture decision to only allow for specifically approved applications, entities, or accesses, with all others being rejected by default. *See, e.g.,* “Information Assurance Division Top 10 Information Assurance Mitigation Strategies,” NSA, (July 2013), *available at* https://www.sans.org/security-resources/IAD_top_10_info_assurance_mitigations.pdf (listing “Application Whitelisting” as the #1 information assurance mitigation strategy.)

[78] Dennis Fisher, “Groundbreaking Cyber Fast Track Research Program Ending,” Threatpost, (Mar. 6, 2013), https://threatpost.com/groundbreaking-cyber-fast-track-research-program-ending-030613/77594/.

adapted to living in polluted soils and are less able to tolerate extreme environmental conditions.”[79] Such voluminous DNA, it turns out, means that the organism is slow to adapt and poorly suited to survive environmental stresses.

More important, however, is that *Paris japonica’s* massive genome is not even biologically necessary.[80] A large genome is often required for more biologically complex species, such as an elephant or a human, (just as a larger codebase is needed for more complex software). But *Paris japonica* is not an unusually complicated flower; rather, it has an unusually large amount of non-coding, or “junk,” DNA.[81] This junk DNA is particularly damaging because it is carries no corresponding benefit in organism complexity. It is, effectively, dead weight.

Although Paris japonica has managed to avoid any ill effects from its genomic dead weight, this is likely a luxury of its stable, non-hostile environment. After all, being poorly suited to adaptation is only a problem if you need to adapt; just as bloated, overly complex software is only really a problem if it is under attack, or needs to be updated. Unfortunately, information security practitioners cannot hope that their poorly minimized systems will simply go unnoticed, as the environment for most information systems is extremely hostile, and the easiest targets tend to get brought down first. Evolution may act randomly, but hackers do not.

The negative effects of large genomes and junk DNA generally are perfect examples of why Minimization is important for information security. Unnecessary complexity is a recurrent source of vulnerabilities in information systems, where bloated source code proves unwieldy and difficult to update, and “junk code” left in codebases opens up unnecessary vectors for attack. Minimization is primarily concerned with removing these unnecessary complexities, allowing for more clear, simple, and streamlined systems and strategies. While Minimization does not mandate removing necessary or functional complexity, it does mandate removing the “junk” that too often permeates security critical systems and strategies.

Strategies & Challenges

Implementation of the Minimization principle can be broken down into two primary stages: (1) *Design* - designing systems to minimize their size and complexity, improving their defensibility and limiting their value; and (2) *Maintenance* - maintaining systems in a streamlined state, by restricting unnecessary growth, resisting unnecessary complexity, and eliminating existing elements when they are no longer relevant or useful. Although both phases are extremely important for robust security, the *Maintenance* phase is likely to comprise the bulk of the security practitioner’s efforts, as security practitioners often have limited input during the design of systems. And even when security is incorporated by design, the longevity of many IT systems means that the practitioner’s predominant activity

[79] Royal Botanic Gardens, Kew, “Rare Japanese plant has largest genome known to science,” ScienceDaily, (Oct. 7, 2010), https://www.sciencedaily.com/releases/2010/10/101007120641.htm
[80] In perhaps the most popular comparison, the human genome is in fact smaller than that of some single-celled protists. This mismatch between complexity and size is attributed to non-coding DNA.
[81] Although recent developments in genomics suggest that this non-coding DNA is not “junk,” we have adopted the simplistic interpretation for illustrative purposes.

will be in maintaining those systems in a minimized state.

With regard to the *design* of systems, Minimization often shares considerable overlap with Compartmentation. Minimizing the functions of system elements necessarily improves Compartmentation, as each minimized element should do one, and only one, thing.[82] Yet Minimization at the element level can also be in conflict with Minimization at the system level, as separating each individual function in some ways adds to the system's overall complexity.[83] Finding the appropriate balance between minimizing overall system complexity and minimizing the complexity of the underlying system elements is an important process in secure system design.

Finally, the practitioner should be keenly aware of unnecessary additions and excessive expansion, falling back on the mantra YAGNI, or "*you aren't gonna need it.*"[84] (Whereas if such additions are necessary, Compartmentation will provide the structure to help make this expansion functional, while still secure.) Considering the difficulty of modifying system architecture and the long lifecycles of some IT decisions, there is a great deal to be saved by doing it right the first time.

Whereas the much more common activity performed by practitioners will be to *maintain* systems in a minimized state. This act of maintenance, much akin to Comprehensivity and Opportunity, is an ongoing process, and requires the practitioner to continually look for opportunities to remove elements as they become unnecessary, to resist new unnecessary additions, and to resist new unnecessary complexity. As systems evolve, and new features are added, old code may become obsolete, old processes may become redundant, and those new features can be implemented in a more secure, more minimized manner. It is the job of the practitioner to be constantly on the lookout for these opportunities, with pruning shears in hand, to maintain the system in a minimized state. Indeed even when approaching a new, relatively static system, there may still be opportunities to streamline functions and reduce the system's overall complexity. This proactive, ongoing approach benefits from Opportunity and from Rigor, as these will help ensure that the practitioner is constantly on the lookout for opportunities to minimize their systems.

It is this maintenance that also likely poses the greatest challenge to practitioners, as resisting growth is in many ways in direct opposition to the natural inclination of information systems, which tend to increase in size, complexity, and interconnectedness. And although important, this desire to resist added complexity is often overcome by the needs of the mission to grow and adapt. As such, effective implementation of Minimization requires a mastery of Proportionality, both for the practitioner to effectively determine when further minimization is warranted, and for the practitioner to be able to convey the competing factors to decisionmakers when arguing for greater minimization.

[82] *See, e.g.,* "separation of concerns" defined as "divide your application into distinct features with as little overlap in functionality as possible. The important factor is minimization of interaction points to achieve high cohesion and low coupling." Key Principles of Software Architecture, Microsoft, https://msdn.microsoft.com/en-us/library/ee658124.aspx#KeyDesignPrinciples.

[83] *See, e.g.,* "least common mechanism" defined as "mechanisms used to access resources should not be shared." Design Principles for Security Mechanisms, InformIT, http://www.informit.com/articles/article.aspx?p=30487&seqNum=2.

[84] YAGNI, or "you aren't gonna need it," is a common mantra in programming that recommends against unnecessary additions to software. This is closely related to KISS, or "keep it simple, stupid," which again cautions against needless complexity in system design.

5: Compartmentation

Isolate system elements, and enable and control the interactions that are strictly necessary for their intended purposes.

"Is this made of distinct parts with limited interactions?"

> *Keep it secret. Keep it safe.*
>
> → **Gandalf, *The Lord of the Rings***

> *We shape our buildings; thereafter they shape us.*
>
> → **Winston Churchill**

Defenders gain an advantage when security and resilience are built into the systems they protect. The Compartmentation principle forms the foundation for this secure architecture: we physically and logically separate systems that have no business commingling; we build those systems in discrete and modular forms; we keep secrets secure by limiting both who has access and how much they have access to; we put sensitive equipment under lock and key; and we architect nested layers of security. Compartmentation is the art of thoughtfully and conditionally breaking up an otherwise connected world: of giving ourselves the opportunity to stem the tide of an attack.

- By placing critical systems on network segments separate from office systems, without the potential for traffic between the two, you insulate those critical systems from attacks brought by machines simply checking email and browsing the web. *Segregate sensitive and/or vulnerable assets from less sensitive or less secure ones to provide barriers to compromise.*
- Using a single encryption key for all of your communications can be risky, as the compromise of that one key will unravel the protections used for all of your past communications. To combat this, practitioners utilize session keys, so the compromise of long-term keys does not compromise past session keys. By compartmenting away separate communications behind separate sessions keys, *practitioners can make their systems "forward secret," and better ensure that the confidentiality of communications is preserved*.[85]
- Minimizing who has access to valuable resources will always be limited by how those resources are structured. If all of your books are stored in the same library, you can't easily restrict access to specific books, because anyone with access to the library can probably reach those restricted books. *Least privilege requires not only that access to resources be limited, but that the structure of those resources allows you to limit access to them.*
- Similarly, implementing Least Privilege without compartmenting those privileges (e.g. separation of privileges) is of little practical use, because a system with only two levels of privilege – user and admin – will end up granting admin status to far more people than it would if more granular levels of privilege existed.
- In software architecture, Compartmentation is essential in preventing certain

[85] For a discussion of the implementation of Forward Secrecy, *see, e.g.*, "Correctly Implementing Forward Secrecy," SANS, (Mar. 13, 2014), https://www.sans.org/reading-room/whitepapers/bestprac/correctly-implementing-secrecy-35842

> classes of abuses directly, but it is even more important in enabling quick and effective patch development, verification, and roll out. *A monolithic design must be edited as a whole; a modular design can be fixed and tested at the point of failure.*

Of all the principles, Compartmentation shares the closest relationship with Minimization. Combined, these two principles form the backbone of secure information architecture, building out a world that advantages the defender. But more fundamentally, Minimization and Compartmentation each require the other for their effective implementation: Compartmentation without Minimization leads to individual “compartments” growing too large or interconnected, undermining the value of Compartmentation; whereas Minimization without Compartmentation leads systems to be small, jumbled messes, undermining the speed and adaptability that Minimization is intended to facilitate. This close relationship often means that Minimization and Compartmentation should be implemented in tandem, applying the steps of Compartmentation to Minimization, and vice versa.

Since these architectural principles often can become confusing in the abstract, consider a common real world example: watertight compartments in seafaring ships.[86] The goal of the watertight compartments is to prevent a single hull breach from sinking the entire ship: even when a compartment is allowed to flood, it keeps water sequestered from the rest of the ship, limiting the impact of the breach. But this structure only works if the compartments are small enough that the flooding of one compartment doesn’t mean you take on so much water that it sinks the ship. A ship with only two watertight compartments is no better off if flooding either compartment still means the ship will sink. In this way, Compartmentation also needs Minimization to make its underlying objectives a reality.[87]

Compartmentation’s architectural nature makes it one of the most impactful principles in terms of practical security outcomes. Compartmentation failures are at the heart of catastrophic breaches,[88] as poorly compartmented systems allow for minor compromises to spread throughout the system unchecked. Additionally, the failure to compartment raises the length and cost of incident response and recovery substantially, as unrestrained interconnectedness can make the extent of a compromise difficult to ascertain. Although the underlying structure of a given system is easy to take for granted, this baseline structure governs how all actions, both friendly and hostile, impact that system, and it is important that practitioner takes the steps necessary to give themself the upper hand.

Grounding

In the first century C.E., Rome was the most populous city the world had ever

[86] *See, e.g.*, Ship Compartmentation and Watertight Integrity, http://www.globalsecurity.org/military/library/policy/navy/nrtc/14057_ppr_ch3.pdf.

[87] However, this logic must always be balanced with Proportionality. At some point, ship Compartments become so small and so numerous that they impede the greater mission of the ship, without the corresponding benefit to security to justify it.

[88] *See, e.g.*, “ransomware.” Tom Risen, “Cisco Reports Ransomware is the Most Profitable Hacker Scam Ever,” US News, (July 27, 2016), http://www.usnews.com/news/articles/2016-07-27/cisco-reports-ransomware-is-the-most-profitable-malware-scam-ever.

seen, housing an estimated 1 million people, and serving as the political, economic, and cultural center of the Western world. Yet Rome lacked the central planning befitting such a grand metropolis, with effectively no building codes, and little overarching structure applied to its broader organization. Worse still, Rome's buildings consisted mainly of shoddy wooden structures, tightly packed together. Unsurprisingly then, Rome was a veritable tinderbox, where small fires could spread rapidly across the entire city. And indeed this is exactly what happened in 64 C.E., when the Great Fire of Rome destroyed two-thirds of the city. (All while the Emperor Nero famously, if anachronistically, fiddled.[89])

Despite the extent of the Great Fire's devastation, Rome was not without firefighting technology: large firefighting forces had been utilized for at least 100 years prior. The firefighters would transport buckets of water to the fire, and would use demolition tools to preemptively destroy buildings in the path of the blaze. (The latter tactic is still used today to fight forest fires.) Yet the crudeness of these methods shows how difficult the firefighters' job was: the city itself was working against the firefighters by aiding the spread of the flames.

Despite happening almost 2000 years ago, the Great Fire of Rome is a perfect example of a Compartmentation failure, where a massively destructive event could have been almost entirely mitigated by using a simple firewall. Not the type of firewall normally thought of in information security, but a literal "fire wall": a flame-resistant structural that divides buildings and cities into distinct compartments to contain the spread of fires.[90] Even if a fire ravages a given building, its firewalls ensure that the fire cannot spread past that compartment. By imposing a baseline of Compartmentation, the Great Fire of Rome could have almost surely been greatly mitigated, if not avoided entirely.

Fittingly, Nero used the massive destruction of the Great Fire as an impetus to instate new building regulations, which mandated the use of fire-resistant brick, (instead of wood), and created space separation (air gap) requirements between buildings, as compared to the prior hodgepodge. By changing the architecture of the city itself, Nero was able to drastically improve the ability to fight the spread of fires, and greatly reduce the impact any given fire would have in the future. While these measures did not completely prevent future fires from breaking out, the process of compartmentalizing[91] the city likely prevented another "great fire" like the one in 64 C.E. Yet information security practitioners need not wait for their "Cyber Great Fire" to begin applying Compartmentation. Just as Nero could have begun compartmenting before nearly all of Rome burned down, so too can practitioners begin Compartmenting their own systems to prevent individual breaches from causing systemic damage.

Strategies & Challenges

[89] Although accounts differ, the closest historical text is that Nero sang and played the Lyre while watching Rome burn. Others sources state that Nero was not in Rome for the Great Fire.

[90] Firewalls serve to segment potential fire hazards between a large range and size of structures, from within buildings, to larger city structures, to tunnels, aircraft, automobiles, and more.

[91] Compartmentalization is a term of art in passive fire protection, and should not be confused with Compartmentation, as used here. Although the concepts share marked similarities, Compartmentation is intended to extend to a much broader range of activities, discussed herein.

As Compartmentation encompases an enormous range of software and system architecture best practices, an exhaustive discussion of the nuances of Compartmentation is beyond the scope of this essay. However, practitioners consistently follow three basic steps when compartmenting valuable or sensitive information or system elements: (1) **isolate** each element on a given system;[92] then (2) **enable** the specific interactions that are needed between elements and between systems; and (3) **control** those enabled interactions that occur between elements and between systems.

> *Before delving into the three steps, however, it is important to note that Compartmentation applies in both the physical and logical sense, and that both physical and logical Compartmentation are important for security. Physical compartmentation, at its most basic, is about erecting walls, limiting the doors through those walls, and putting guards at those doors. Logical compartmentation, by contrast, is about building systems in isolation, limiting the protocols that allow those systems to interact, and monitoring and controlling the traffic that occurs via those remaining interactions. But in practice, most systems require both physical and logical compartmentation, (and the physical/logical split tends to bleed together in practice, as with locks on doors). As such, it may be necessary to go through each step twice, first considering the application to physical systems, and another for the application to logical systems.*

The first step, *isolate*, requires that you architect systems and elements so that they are discrete, non-overlapping, and conceptually distinct from one another. Put simply, you want to divide your system into separate conceptual boxes, each of which accomplishes one, and (ideally) only one thing. The authentication server should only perform authentication; the safe should only store valuables. By isolating, you are fighting against the unrestrained interconnectivity that pervades the modern world, and defining your systems in terms that are discrete and limited.[93] Although articulating the *isolate* step often feels like common sense, ("what else would you put in the safe?") the world is replete with examples of vulnerabilities arising from individual systems or elements trying to do too much.[94] As such, many information security best practices reflect this need to isolate, with "modular design" and "separation of concerns" recommending dividing up tasks, and "separation of privileges" recommending dividing up how accesses are distributed.

> *It is important to note, however, that "isolated" elements may in fact be very close, both physically and logically. For instance, a great example of Compartmentation is a safety deposit box: although each safety deposit box is physically very close, they are isolated from one another, both physically (by metal barriers), and logically (by dual-key locks).*

[92] Recall that systems themselves can be elements of larger systems, and that their elements may themselves be systems, such that Compartmentation applies up and down the scale of system complexity.

[93] The isolate step is therefore reflected in the transition from blacklisting to whitelisting. Blacklisting starts from a position of interconnectivity, and attempts to limit and control all traffic on an ad-hoc basis. Whitelisting, by contrast, begins with a world that is isolated, and therefore is empowered to choose the limited traffic that it will enable and control.

[94] For a common example, the concept of "choking" in human anatomy is largely a function of a failure to isolate. The gastrointestinal and respiratory systems did not need to utilize the same mechanisms: specifically, the throat. Choking therefore is a form of human vulnerability that better Compartmentation could have eliminated almost entirely.

> *Moveover, the safety deposit box shows that an "isolated" element may in fact be nested within another element of the same system, since the individual safety deposit boxes are all encompassed within a single bank vault. Despite the apparent connectedness of this system, no individual key can compromise the entire system, because the elements are fundamentally isolated.*

The second step, *enable*, follows closely from the first, by acknowledging that modern systems require some interconnectivity to complete their missions, but limits the interactions that connect those systems to only those which are specifically enabled. Effectively, Compartmentation shifts the assumption from defaulting toward more connectivity, to default toward more isolation. After all, Compartmentation is not intended to impede interconnectedness that is necessary to advance the mission, but it also understands that interconnectivity can pose a significant threat to the mission.[95] The more interconnected a system or element is, the greater the risk it poses if compromised. To this end, Compartmentation seeks to limit the interactions for each system and element to only those that are strictly necessary to further their mission, commonly referred to as least privilege.

The *enable* step in many ways reflects the close interplay between Compartmentation and Minimization. Just as Minimization recommends least privilege as a means of reducing the overall attack surface of a given system, so too does Compartmentation recommend least privilege to mitigate the fallout from the compromise of any individual element in that system. And perhaps more importantly, least privilege is of little use if the underlying privileges aren't compartmented. A system with only two levels of privilege, e.g. user and admin, may technically adhere to least privilege, but will still allow far more access than is necessary. Alternatively, a system that imposes least privilege without compartmenting system functions, (i.e. *separation of concerns)*, may lead to a scenario where widespread access to a general resource leads to unnecessary access to an improperly connected resource.

The final step, *control*, is about managing the interactions arising from the *enable* step. Even when certain interactions between systems and elements are required, they must still be implemented in a manner that allows for them to be controlled. Doors must have locks, and those locks should be changeable; hallways should have cameras, and alarms when unauthorized persons gain entry. In the information security world, this means that inputs and outputs should be scrubbed, logs should be taken and audited, and procedures for revoking credentials should be put in place. In all cases, the importance of the *control* step is in ensuring that the practitioner can take meaningful action over the interactions between the systems and elements they are tasked with designing, implementing, and

[95] Because of the importance of interconnectivity in the modern world, Compartmentation requires a careful Proportionality assessment to determine the appropriate balance between interconnectivity and Compartmentation. So, for instance, the Northeast Blackout of 2003, in which over 55 million people lost power in the United States and Canada, should not be viewed purely as a Compartmentation failure. As it happened, part of the design of the electrical grid relied on interconnectivity, so that if an individual node failed, other nodes could pick up the slack. However, this interconnectivity also created the potential for cascading failure, where the excess load from failed nodes continually overwhelms the remaining nodes. Although this problem could have been solved via better Compartmentation, this might have undermined one of the goals of the mission - ensuring constant power to as many people as possible.

operating, to protect and advance the needs of the mission.

Too often, however, control is the only step of Compartmentation that practitioners are empowered to employ. While the first two steps of Compartmentation are architectural, and therefore difficult to retrofit onto existing systems, control is more easily put in place after-the-fact, at least to some extent.[96] Yet without the rest of Compartmentation, *control* becomes an exercise in futility, where practitioners are tasked with overseeing networks and systems that are sprawling messes of interconnectivity. It is difficult enough to keep track of someone or something's activities when you don't know what all they are empowered to do, and harder still to stop them from acting altogether. Although it is certainly possible to exercise control in the absence of this underlying structure, that control will be fundamentally hampered when the system architecture is not designed to facilitate control.

Although the benefits of Compartmentation are clear, effectively implementing Compartmentation can prove challenging. To begin, Compartmentation typically needs to be incorporated during the design phase of a system to be effectively implemented, a time when system functionality is still malleable.[97] Making substantial changes to existing system architecture is often infeasible, and the costs of redesigning these legacy systems often outweigh the risks they pose. Indeed it took a "great fire" for Nero to rebuild Rome. To address Compartmentation in these ongoing scenarios, the practitioner will likely need to adopt a strategy similar to that in Minimization, where the practitioner attempts to better structure these legacy systems in an ongoing and continuous manner.

Furthermore, Compartmentation requires a constant and close examination under Proportionality, as over-compartmenting a system can greatly increase costs, inflate system complexity, and clash with usability and the benefits of connectivity. Indeed shoddy or overzealous implementation of Compartmentation can in fact increase the risks facing a system, particularly if the complexity of that system outstripes the capabilities of the security team to develop and manage it effectively. Nevertheless, it is the job of the practitioner to strike these balances, and make a determination of how best to compartment their systems to further the needs of the mission.[98]

[96] For instance, it is typically very easy to add visibility to an existing system, like putting a camera in a hallway. However, without the rest of Compartmentation, there will be far too many hallways, and the intended uses of those hallways will be complex and varied, making the practitioner's job substantially more difficult, as they will have difficulty determining when and whether problematic behavior is occurring.

[97] However, the importance of "security by design" has been given a greater degree of attention in recent years, as stakeholders have become increasingly aware of the difficulties in retrofitting security onto existing systems.

[98] This may be through better hiding Compartmentation behind technical solutions, performing risk assessments to determine whether connecting two separate systems is acceptable, or implementing response mechanisms that can isolate interconnected systems in the case of a breach.

6: Fault Tolerance

Anticipate and address the potential compromise and failure of system elements and security controls.

"What happens if this fails?"

> *Such a strategist was the king that he had a contingency plan for his contingency plan, and even, if circumstances required, a contingency plan for his contingency plan's contingency plan.*
>
> → **Frank Beddor, *Seeing Redd***

> *If one thing goes wrong, everything else will, and at the same time.*
>
> → **Peter Drucker**

In an uncertain world, failures are all but inevitable. Whether due to the laws of thermodynamics, laws of complexity,[99] Murphy's law,[100] or a wily adversary, it is best to assume the worst, and plan accordingly. Any person can make a mistake or become unavailable; any piece of technology can have an undiscovered flaw or need maintenance; any process can fail in design or implementation. But to allow any of these to become a 'single point of failure' would create a systemic weakness, because these failures are likely to occur sooner or later. Fault Tolerance teaches practitioners to anticipate these failures before they occur, and to take steps to mitigate or eliminate the fallout from those failures preemptively. Whether by layering defenses, compartmenting assets, switching to more robust controls, or speeding response and recovery, Fault Tolerance allows for systems to remain resilient in an uncertain world.[101]

- Any single defense can fail completely, or be interrupted or circumvented. Interruptions and circumventions do not have to be total failures if additional defenses are present to catch intrusions that make it past the first line. *Practitioners should adopt a defense in depth strategy that uses multiple layers of defense to insulate against individual breaches.*
- Authentication credentials like encryption keys, credit card numbers, and passwords are frequently compromised and quickly disseminated. Rather than allow for a single breach to permanently compromise a system, *credentials should be easily revocable and processes should be put in place to detect these compromises and replace credentials quickly in response.*
- If only the CISO can authorize incident response, the CISO being unavailable will cause incident response to grind to a halt. *There must be alternates for key players in all mission-critical processes.*
- Procedures are just as susceptible to failures and compromises as systems, and

[99] Complexity science is an emerging interdisciplinary field of study which analyzes and models complex systems. A chief component of complex systems is that they generate emergent properties which would not be discerned by looking solely at the component parts. For a discussion of complexity science and cybersecurity, *see, e.g.*, Burns, *supra*.

[100] Murphy's law is a popular adage which states that "Anything that can go wrong, will go wrong."

[101] *See also* Peter A. Lee & Thomas Anderson, "Fault Tolerance: Principles and Practices," Springer, (1990), DOI:10.1007/978-3-7091-8990-0.

procedures that are too rigid or rote can be easily undermined or exploited. *Security procedures should strive towards adaptability, operating as a security immune system that can quickly respond to new and emerging threats, and update based on changes in their environment.*

- For mission critical systems, ensuring that some functionality remains, even in the case of catastrophic failings, is often a necessary requirement. *Design mission critical systems to ensure survivability of their most essential features even under extreme failure conditions.*[102]

Fault Tolerance is the single most important principle for improving resilience. Resilience is a security objective[103] characterized by continued operability in the face of attack, as well as the ability to respond and adapt to hostile environments.[104] As such, true resilience requires the use of all of the Principles in combination.[105] Nevertheless, Fault Tolerance can be considered the primary principle that drives resilience, by forcing practitioners to consider the impact of failures and advanced adversaries on their ability to achieve their security goals. By anticipating and addressing the failures that are likely to occur in hostile environments, practitioners can take steps to ensure that their critical systems and strategies continue to operate, and that they are able to adapt and respond to better face these challenges in the future.

The importance of Fault Tolerance has never been more pronounced than in the contemporary world, where operational continuity, resilience, and breach containment are often more important than outright prevention. Indeed, even mission-critical technologies are frequently susceptible to attack, and assuring continued operability for these critical systems requires the adoption of a Fault Tolerance approach. And outside of critical systems, much of our economy is reliant upon continuous access and operability, from e-commerce to social media to advertising-driven businesses generally.[106] These organizations are forced to go beyond mere prevention, and design systems under the assumption that their security controls will fail.[107] While such breaches are never desired,

[102] Knight et al., Towards a Rigorous Definition of Information System Survivability, UVA, (2003) http://www.cs.virginia.edu/~jck/publications/discex.2003.pdf.

[103] We view resilience as analogous to the CIA triad: security objectives which we strive to achieve. Resilience does not guide decision-making directly, and as such does not fit our requirements for a principle.

[104] "Resilience," however, is often subject to broad and competing definitions. For our purposes, we will be adopting the definition put forth in Presidential Policy Directive 21, defining resilience as "the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents."

[105] For instance, if you have more moving parts than you can manage (failure of Minimization), you are not resilient. If you don't understand what you have or what is happening to it so that you can respond in real time (failure of Comprehensivity), you are not resilient. If your organization, environment, and systems are not each in turn bounded into comprehensible parts with clearly-defined interactions (failure of Compartmentation), you cannot contain damage or understand the spread and impact of a compromise and are not resilient. If you've created a single point that can compromise everything (failure of Fault Tolerance), you are not resilient.

[106] *See, e.g.*, Areil Tsietlin, "How Netflix Embraces Failure to Improve Resilience and Maximize Availability, Usenix," *available at* https://www.usenix.org/conference/lisa13/how-netflix-embraces-failure-improve-resilience-maximize-availability.

[107] "Making Your Enterprise Cyber-Resilient," Accenture, (May 1, 2016),

organizations can take steps to enhance resilience and ensure that critical functions will continue to operate notwithstanding.

Fault Tolerance is most important to understood in conjunction with Compartmentation: Compartmentation is the architectural practice that makes system elements discrete and isolatable, and therefore the failure of those elements easily understood and analyzed. Fault Tolerance is the strategy that capitalizes upon Compartmentation's design to anticipate and address the failures of each compartment. Without Compartmentation, Fault Tolerance is extremely difficult to meaningfully implement, because you cannot prepare for failures when you do not fully understand both how those failures might occur and what the fallout from a given failure will be. In well-compartmented systems, the fallout from the failure of any individual element should be easy to determine, as the only interactions with that element are specifically enabled, and their traffic controlled.

However, Fault Tolerance goes beyond this, and must also be understood in combination with Comprehensivity and with Rigor. Comprehensivity teaches that failures occur not just across space, but across time, and Fault Tolerance must be prepared to address failures that occur far in the future, such as through *patching*. Whereas Rigor is important in providing mechanisms by which failures can be quickly and efficiently addressed, by building in processes that can handle failures without breaking down, such as through *secure exception handling*.[108]

Grounding

On January 28, 1986, the world was stunned and saddened by the Space Shuttle Challenger's sudden explosion 73 seconds after initial launch. In the aftermath of the disaster, President Ronald Reagan initiated a Presidential Commission to determine what exactly went wrong. The findings from this inquiry, collected in the Rogers Commission Report,[109] with special commentary by Richard Feynman,[110] revealed a systemic failure by the US National Aeronautics and Space Administration (NASA) to recognize, appreciate, and respond to signs of component failures in the Challenger's design. Rather than anticipate and address potential failures, NASA proved to be in denial as to the real risks the mission faced, resulting in calamitous ends.

The Rogers Commission Report and Feynman's essay identified a pattern within NASA management of mischaracterizing signs of potential failure as signs of Fault Tolerance. For example, the proximate cause of the Challenger disaster was the faulty O-rings on the solid rocket booster. Upon further investigation, the Rogers Commission discovered that signs of O-ring deterioration had been discovered and highlighted as a potential threat months

https://www.accenture.com/t20160501T215418__w__/us-en/_acnmedia/PDF-15/Accenture-Making-Your-Enterprise-Cyber-Resilient-v2.pdf.

[108] *See, e.g.*, Aslan Askarov and Andrei Sabelfeld, "Catch me if you can: permissive yet secure error handling." PLAS 09, 45-57. DOI=http://dx.doi.org/10.1145/1554339.1554346.

[109] Rogers et al., "Report of the Presidential Commission on the Space Shuttle Challenger Accident," (1986), *available at* https://history.nasa.gov/rogersrep/v1ch6.htm. ("Rogers Commission Report.")

[110] Richard Feynman, "Personal observations on the reliability of the Shuttle," (1986), *available at* https://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/Appendix-F.txt.

prior to launch, but that this worrying sign was largely disregarded. Rather than view the O-Ring's deterioration as a threat to the mission, management emphasized that the deterioration was only partial, and therefore a sign of the component's resilience. Put another way, NASA effectively said *the O-rings only deteriorated one third, this proves they are safe,* rather than recognize that this deterioration was itself a sign of component failings that might become much worse.

Although the O-ring deterioration proved to be the most salient example, the Challenger disaster highlighted a widespread practice of disregarding, misconstruing, or otherwise failing to address signs of failure among NASA management. Indeed when asked what they believed to be the likelihood of individual component failures, management would often give answers that were orders of magnitude smaller than that of their engineers. Rather than embrace a mindset of Fault Tolerance, NASA management remained in denial as to the risks they faced, making them, unfortunately, too tolerant of faults.

Information security practitioners can learn a great deal from the Challenger disaster, particularly with regard to the importance of understanding, anticipating, and addressing potential signs of failure. Fault Tolerance is about rejecting the ease of denial, and taking a step toward harsh realities: failures will happen, and you must be ready to address them. Rather than assume the best, practitioners should assume the worst, so that even if the worst possible scenario comes to pass, their systems will still provide acceptable outcomes.

Strategies & Challenges

Implementation of Fault Tolerance, put simply, requires the practitioner to ask the questions "What happens when this fails?" and "What happens when this is compromised?". Although seemingly straightforward, this can prove exceptionally difficult in practice. To aid in this difficulty, there are two primary strategies for implementing Fault Tolerance at a practical level: those that start by analyzing high-level failures and compromises, and work backwards to assess how to prevent them ("top-down" approaches);[111] and those that assume failures or compromises at the element or control level, and assess the impact on the larger system or strategy ("bottom up" approaches).[112]

> *Although the methodology for these two approaches is largely drawn from non-adversarial expertises, such as safety engineering, the structure of the analysis remains largely unchanged in an adversarial context. In either case, the ultimate concerns are relatively objective and therefore assessable before the fact. While the introduction of adversarial actors greatly increases the scope of potential causes of failure, and may make identification difficult, the methodology for analysis is largely unchanged. As such, we will include "compromises" in with "failures" for the remainder of*

[111] *See, e.g.,* "Fault Tree Analysis," IEC 61025, https://webstore.iec.ch/preview/info_iec61025%7Bed2.0%7Den_d.pdf.

[112] *See, e.g.,* "Fail Mode and Effect Analysis," Weibull, http://www.weibull.com/basics/fmea.htm. *See also* "Fail mode, effects, and criticality analysis," (FMECA) which includes an assessment of the criticality of each failure.

this essay.

In “top-down” Fault Tolerance strategies, practitioners start with a select number of well-defined, high-level “failures” which they are specifically attempting to anticipate and address. For instance, an e-commerce site may identify the site being offline as a high-level failure, and then work backwards to determine how to prevent this from occurring. Top-down strategies are particularly useful for Fault Tolerance where the failures are clearly identified, for systems at high levels of abstraction, or for systems with large amounts of human interaction, and for systems where more in-depth Fault Tolerance strategies wouldn’t be proportionate to the risk. Top-down approaches are also particularly useful in adversarial contexts when the defender has specific threat intelligence regarding the means and methods of their adversaries. By utilizing a top-down approach, practitioners can ensure that Fault Tolerance is clearly targeted at the failures that will have the most direct and substantial impact on the mission, and take the steps necessary to address those contingencies.

Whereas in “bottom-up” Fault Tolerance strategies, practitioners take a much more comprehensive approach to analyzing system failures, assessing the failure of each of the system’s elements in an iterative manner.[113] Bottom-up approaches assume the failure of each element in their system, and analyze the impact of that failure on the system as a whole. Bottom-up approaches are particularly useful in mission-critical systems, systems where the potential failures are unknown, systems still in development, and systems that are expected to undergo significant environmental stress. Bottom-up approaches present many advantages over top-down approaches for modern information systems, as often the failures that will have the greatest impact on the mission aren’t easily discovered beforehand, or may arise from unexpected or unconventional mechanisms. However, bottom-up approaches are also more burdensome to implement, and create the potential of “missing the forest for the trees.” As such, the choice between bottom-up and top-down approaches will vary depending on the needs of the mission, the system to be protected, and the resources available.

Despite this conceptual divide, however, in practice practitioners can and should utilize both approaches, to varying degrees, allowing for security that focuses on high-priority, high-level failures, while also assessing systems comprehensively at the element level.[114] Pure top-down and bottom-up approaches both present notable shortcomings, and strategies that utilize both approaches can minimize these shortcomings while maximizing the value to the mission. So, for instance, an organization may choose to utilize top-down approaches when analyzing systems of sufficiently great complexity, such as the entire organization, while implementing bottom-up approaches as the complexity of the system decreases, and in particular, during development. And since both strategies operate in a hierarchical manner, the benefits of bottom-up analyses will cascade upwards, while the

[113] For an closely analogous concept, in the Feynman Report, Richard Feynman discusses bottom-up design, a closely related concept to bottom-up fault tolerance, wherein products are developed at the component level first, allowing faults or incompatibilities to be addressed without threatening some preimposed overarching plan.

[114] As with all Principles, Proportionality tempers what might otherwise seem to be an overwhelming task in implementing Fault Tolerance. Unrestrained Fault Tolerance is theoretically infinite, as you can always add more redundant backups, or more layers of defense.

benefits of top-down approaches will percolate downwards.

Finally, once these points of failure have been identified, the practitioner must determine the appropriate remedy to address those contingencies. While an exhaustive list of Fault Tolerance remedies is beyond the scope of this essay, most remedies will fall into one of a few categories:

1. Improving redundancy (e.g. backups);
2. Improving survivability (e.g. partial operability);
3. Increasing defenses for indispensable elements;
4. Speeding response and recovery;
5. Greater reliance on other Principles; and
6. Outright replacement.[115]

The last strategy, replacement, warrants some further comment. Although it may be tempting to think of Fault Tolerance purely as the principle of “have a Plan B,” in many cases Fault Tolerance requires practitioners to “get a better Plan A.” After all, often the best way to “address” potential failures is to switch to a system or strategy that isn’t susceptible to that specific failure to begin with. So while Fault Tolerance is primarily concerned with contingency planning, any solution that removes that contingency entirely is still adhering to Fault Tolerance.

[115] In all cases, the need for Fault Tolerance remedies will be balanced by the Proportionality principle.

7: Proportionality

Tailor security strategies to the magnitude of the risks, accounting for the practical constraints imposed by the mission and the environment.

"Is this worth it?"

> *If the highest aim of a captain were to preserve his ship, he would keep it in port forever.*
>
> → **Thomas Aquinas**

Anything worth doing involves taking risks. Proportionality reflects a basic compromise familiar to all security practitioners: security does not exist in a vacuum, and must be balanced against the competing interests that security impedes. The front door and the bathroom door rarely require the same protections, so security for each should be appropriately tailored. Security costs money, time, and resources, and it (rather intentionally) gets in the way. While it is tempting to assume that a perfect world would always feature perfect security, in practice, this would be costly, unwieldy, and bad for business. Practitioners must be able to acknowledge the costs security imposes and effectively weigh those limitations against the risks presented.

- Security should not exist solely for the sake of security. Security is always in furtherance of some greater organizational goals, and it should be crafted with those goals in mind. Rather than simply implement security in a haphazard manner, *security should be targeted towards the needs of the business*.[116]
- Spending $1000 to secure $100 is not good security, even if that money is never stolen. Whereas spending $10 to secure that same $100 doesn't become bad security just because it was stolen. *Good security should be tailored to the risk, and shouldn't be evaluated by individual security outcomes.*
- Determining whether security is having a meaningful impact is a constant challenge, and makes evaluating potential security additions difficult to quantify. To confront this problem, *organizations should adopt a risk-management approach to allow for greater strategic decision-making about the risks facing their organization, and to incorporate security into broader organizational strategies*.[117]
- Security controls often create substantial burdens on organizations through basic usability, quality of life, and financial costs, and may offer few comparative benefits for security. *Removing superfluous security can improve usability, cut costs, and streamline business processes.*[118]

[116] *See, e.g.*, Principles for Information Security Practitioners, Information Security Forum, (2010), https://www.isaca.org/Knowledge-Center/Standards/Documents/Principles-for-Info-Sec-Practitioners-poster.pdf.

[117] *See, e.g.*, "Framework for Improving Critical Infrastructure Cybersecurity," NIST, (2014), http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214.pdf. *See also*, "Guide for Applying the Risk Management Framework to Federal Information Systems; A Security Lifecycle Approach," NIST SP 800-30 Rev. 1, (June 5, 2014), http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r1.pdf.

[118] *See, e.g.*, Matthew Rosenquist, "Never Forsake the Fundamentals of Security Part 1," Intel, (Aug. 25, 2014), https://itpeernetwork.intel.com/never-forsake-the-fundamentals-of-security-part-1/.

In each of these cases, the specific practice is balancing the need for security with the competing interests that security impacts. Although Proportionality is often brought up as a way to reduce excessive security, it is important to emphasize that this balancing act goes both ways: scenarios where the risks are high should be afforded the corresponding heightened security.[119] Proportionality works as a tool both to resist overzealous security and to bolster lackluster security. As such, Proportionality is about balance: security should not outweigh its competing interests, nor should they outweigh it. In all cases, the practitioner must strike the appropriate balance.

Because Proportionality emphasizes balance, it serves as the primary limiting factor on the other Principles, and is important across the board for making the Principles workable on a practical level. For instance, although Comprehensivity teaches that security must identify and account for *all* risks, Proportionality teaches that this comprehensive security must also be proportionate. True Comprehensivity is typically infeasible, so Comprehensivity relies on Proportionality to find a happy medium.[120] While ideally this will coalesce towards at least some security across the board, in practice, this may necessitate prioritizing some risks over others. Similarly, a strict reading of Fault Tolerance would lead to infinite redundant defenses for each system, as each backup would require its own backup. Proportionality seeks to find an appropriate middle ground that uses Fault Tolerance in a manner that maximizes the benefit to the mission.

Proportionality is also important for evaluating which Principles are particularly important for a given mission, and developing strategies that accommodate these needs. For organizations where continuous operability and resilience are of primary importance, Fault Tolerance and Compartmentation should take center-stage; whereas for organizations seeking to respond more quickly to changes in the environment, Minimization and Opportunity may be given greater weight. Although all of the Principles are important when crafting a balanced security policy, understanding and applying how specific Principles can be emphasized to further organizational goals should be a primary concern of the practitioner.

Finally, it is important to emphasize that Proportionality should never be understood as condoning the development of security blind spots, nor to encourage recklessness in decision-making. Proportionality does not negate the fundamental lessons of the Principles, nor does it justify embracing reckless or highly speculative opportunities. Proportionality is about striking the appropriate balance when implementing the Principles, and should never be viewed as a tool for disregarding them completely.

With regard to impact, Proportionality is unique among the Principles in that it is primarily reflected by the magnitude of security outcomes. Failures of Proportionality are just that: disproportionate. As such, these failures can range from almost imperceptible to near-catastrophic. An over-allocation of resources may result in the exploitation of other, less-defended vulnerabilities, (or simply to unnecessarily high costs). Whereas an under-allocation of resources can lead to a massive breach through the proverbial "weak

[119] *See, e.g.,* "Five Essential Steps to Improve Cybersecurity," Deloitte, *available at* http://www2.deloitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-risk-cyber-5-steps.pdf.
[120] Indeed, the combination of Comprehensivity and Proportionality can be thought of as setting the scope and magnitude of all security decisions.

link." Indeed in some cases Proportionality failures will lead to little more than the increased hassle of using unnecessary security. Whereas Proportionality successes are defined by security operating efficiently: anyone can achieve good security by spending three times what is necessary, a proportionate solution should be tailored just right. And somewhat counterintuitively, getting Proportionality "right" will sometimes still involve a breach. A single properly balanced defense can still be defeated, but in the aggregate these balanced defenses will lead to best outcomes.[121] A single winning lotto ticket does not make buying lotto tickets good investment advice, and a single breach does not mean security overall was inadequate.

Grounding

During the early years of the Revolutionary War, the Continental army faced a British weapon far more deadly than muskets: smallpox. Unlike their European adversaries, (whose frequent exposure to smallpox and widespread use of inoculation widely protected against the disease), American soldiers had had little direct contact with smallpox, and thus no pressing need to inoculate. This all changed with the outbreak of the war, as British soldiers carried smallpox with them, and even sent diseased men to infiltrate Continental cities. The threat posed by the disease was immediately apparent, and the Continental army used every step available to quell the disease's spread, including quarantines and special commands of entirely smallpox immune soldiers. Every step, that is, except for inoculation.

Inoculation, although seemingly the most obvious solution, presented many downsides. The idea of inoculating up to three quarters of the Continental Army was a logistical feat never before undertaken; the inoculation process was dangerous for the soldier, and could potentially spread to nearby civilians; and, perhaps most importantly, the time required to implement would provide a strategic weakness the British could exploit. As it happened, inoculation largely incapacitated the soldier for several weeks, and this thinning of the Continental forces would provide a tremendous opportunity for the British forces, should they become aware of it. The risks presented were so great that General George Washington initially forbade any inoculation of soldiers, and the Continental Congress banned the practice.

Yet as the war raged on, and the threat and impact of smallpox became more apparent, General George Washington ultimately decided that the benefits of inoculation outweighed the risks. Notwithstanding the practical difficulties, the strategic vulnerabilities, the unpopularity, and indeed even the technical illegality of the program, Washington determined that inoculation was necessary for the continuation of the war, and his decision to do so is claimed by some to be his single most impactful decision during the Revolutionary War.

[121] Exactly how much money to allocate to a given risk has been studied in the academic literature, with some recommendations stating that no more than 37% of the value of the assets being protected should be allocated towards security for those assets, and in many instances, substantially less. *See, e.g.*, Lawrence A. Gordon & Martin P. Loeb, "The economics of information security investment," ACM Transactions on Information and System Security (TISSEC), v.5 n.4, p.438-457, (November 2002) [doi>10.1145/581271.581274].

Information security practitioners can learn a great deal from Washington's dilemma. Security, much like public health initiatives, is often burdensome, unpopular, costly, and uncertain. Yet these countervailing forces do not negate the underlying needs for that security. Washington understood that inoculation provided both major benefits and substantial drawbacks, and was constantly engaged in an internal debate over where to strike the balance. While Washington's decision was binary, most security decisions offer more flexibility, and Proportionality's goal is to strike the balance that best advances the needs of the mission.

Strategies and Challenges

Implementation of Proportionality is focused on balancing security's competing interests: (1) the *risks* present; (2) the *security* offered, (3) the security's *usability*; and (4) the security's *costs*.[122] From the Mom-and-Pop shop – low risk, low security, high usability, and low cost – to the nation's most valuable assets – high risk, high security, low usability, and high cost – the determination of how to protect both of these requires balancing the same types of interests. Even low-risk operations require some security, whereas the most high-risk operations will still curtail security to allow for some usability, and because more security would offer only diminishing returns.

Implementation of Proportionality, therefore, requires a keen understanding of the practitioner's mission, and the ability to weigh how security impacts the mission's goals. Proportionality requires practitioners to take a more holistic approach to security, focusing not on security for security's sake, but on how security furthers the mission. While this is not to suggest that practitioners should completely disregard existing chain-of-command, organizational hierarchies, or the specifics of the task they are given, their actions should nevertheless be informed by how their specific task or role furthers the broader mission.[123]

Although conceptually simple, Proportionality is perhaps the most difficult Principle to "get right" in practice. The four factors that Proportionality balances – risk, security, usability, and cost – each proves to be far less concrete than it first appears. In many fields, the risk a particular system presents is extremely difficult to quantify, impacting intangibles like privacy,[124] reputation,[125] competitive advantage,[126] and future vulnerabilities. Whereas usability is often given disproportionate weight by other organizational elements that don't

[122] Although "costs" is primarily concerned with monetary costs, this can also include less tangible costs, such as societal costs. *See, e.g.,* the debate over full-disk encryption between Apple and the FBI, Arjun Kharpul, "Apple vs FBI: All you need to know," CNBC, (Mar. 29, 2016), http://www.cnbc.com/2016/03/29/apple-vs-fbi-all-you-need-to-know.html.

[123] *See, e.g.,* Maj. Richard Dempsey and Maj. Jonathan M. Chavous, "Commander's Intent and Concept of Operations," Military Review, (Nov-Dec 2013), *available at* http://usacac.army.mil/CAC2/MilitaryReview/Archives/English/MilitaryReview_20131231_art011.pdf.

[124] *See, e.g.,* "2016 Cost of a Data Breach," IBM, http://www-03.ibm.com/security/data-breach/.

[125] *See, e.g.,* "Reputation Impact of a Data Breach," Ponemon, (Nov. 2011), https://www.experian.com/assets/data-breach/white-papers/reputation-study.pdf.

[126] *See, e.g.*, Jeremy Seth Davis, "FDA: Hackers want pharma data for competitive advantage," SC Magazine, (Mar. 16, 2016), http://www.scmagazine.com/fda-hackers-want-pharma-data-for-competitive-advantage/article/496766/.

regularly confront security problems directly, and which tend to view security as mostly an unnecessary hassle. Indeed even the amount of security offered by a given control or strategy can be difficult to identify, as some may be poorly targeted, while others may introduce new vulnerabilities, and all can be potentially undermined by non-comprehensive strategies. And rather ironically, the final factor, cost, although almost certainly the easiest to quantify, is also the factor over which practitioners have the least direct control.

Despite this difficulty, the proper response when implementing Proportionality will nevertheless require mastery of the Principles. By contextualizing security problems in terms of fundamental principles, the practitioner will be empowered to persuade decisionmakers as to why a given strategy or control will be effective, why it will support the mission, and why it is worth the tradeoffs. Good security is good for business, and it is the responsibility of the practitioner to clearly and succinctly explicate the risk, and why the costs are therefore justified.