3-23-2018

Cybersecurity Assessment and Mitigation Stochastic Model

Matthew W. Davis

CYBERSECURITY ASSESSMENT AND MITIGATION STOCHASTIC MODEL

THESIS

Matthew W. Davis, Captain, USAF

AFIT-ENV-MS-18-M-194

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-18-M-194

CYBERSECURITY ASSESSMENT AND MITIGATION STOCHASTIC MODEL

THESIS

Presented to the Faculty

Department of Systems Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Systems Engineering

Matthew W. Davis, BS

Captain, USAF

March 2018

AFIT-ENV-MS-18-M-194

CYBERSECURITY ASSESSMENT AND MITIGATION STOCHASTIC MODEL

Matthew W. Davis, BS

Captain, USAF

Committee Membership:

Lt Col Logan O. Mailloux, PhD
Chair

Dr. Brent T. Langhals
Member

Mr. Harrell Van Norman
Member

Abstract

With numerous cybersecurity incidents and vulnerability concerns in an increasingly contested cyber warfighting environment, the Department of Defense (DoD) has mandated cybersecurity assessment and authorization of all major weapon systems (MWS) before their use. In response to this direction, the Air Force Life Cycle Management Center (AFLCMC) created the Platform Information Technology Assessment and Authorization (PIT A&A) Process. Modeled after the NIST Risk Management Framework (RMF), this process applies a risk-based approach to cybersecurity with the goal of identifying risks and mitigating vulnerabilities in MWS. Within this work, a stochastic model of the PIT A&A Process is presented with an emphasis on understanding how the complexity of systems, accuracy of security artifacts, and workforce proficiency impacts the ability to effectively mitigate cybersecurity risks.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Lt Col Logan Mailloux for his guidance and support throughout the course of this thesis effort. The insight, experience, and patience were certainly appreciated. I would, also, like to thank my sponsor from the Air Force Life Cycle Management Center for the support provided to me in this endeavor.

Capt Matthew W. Davis

Table of Contents

Page

List of Figures

Page

List of Tables

Page

CYBERSECURITY ASSESSMENT AND MITIGATION STOCHASTIC MODEL

I. Introduction

General Issue

On Friday, 12 May 2017, the onset of a major cyber attack known as the WannaCry Ransomware commenced [1]. This attack targeted computers running Microsoft Windows and locked operators out of their data unless they paid a fee. Within one day, more than 2 million computers in over 150 countries reported the infection. Companies such as FedEx, Deutsche Bahn, and the United Kingdom's National Health Service experienced setbacks as a result of this attack. While it can wreak havoc on the commercial sector, the cybersecurity threat is more than just ransomware or traditional network intrusion; cybersecurity is one of the most serious challenges the nation faces [2]. The more heavily reliant a system is on cyber-dependent technologies (e.g., software, interconnectivity, etc.) the more vulnerable it is against cyber attacks. The US DoD's ability to successfully execute mission cybersecurity largely depends on the competitive advantage gained by leveraging advanced cyber systems. When asked how frequently DoD systems are under attack, Commander of U.S. Cyber Command (USCYBERCOM), General Keith Alexander, reported that U.S. military networks experience "hundreds of thousands of probes a day" [3]. Thus, the cyber threat is real and persistent, causing the DoD to question whether critical weapon systems will function as expected when called upon [4].

This concern is so significant that it prompted a recent congressional mandate for the DoD to assess all Major Weapon Systems (MWS) for cyber vulnerabilities [5]. The

Air Force's response to this mandate included standing up a brand new capability - the Cybersecurity and Resiliency Office for Weapon Systems (CROWS) in 2017 [6]. The need to stand up the CROWS is evidence that the DoD's current process of developing cyber-secure systems requires re-examination.

Background

In the 1990s, the assessment and certification of information systems within the DoD were accomplished through the DoD Information Technology Security Certification and Accreditation Process (DITSCAP), which was transformed into the DoD Information Assurance Certification and Accreditation Process (DIACAP) in the late 2000s [7]. More recently, the DoD acknowledged the need to adapt to an increasingly dynamic threat environment and moved to the National Institute of Standards (NIST) risk-based approach known as the Risk Management Framework (RMF) [8]. The RMF provides mission owners and supporting organizations the ability to identify and prioritize risks in order to implement appropriate mitigations with the goal of reducing and (ideally) eliminating critical vulnerabilities. The RMF provides a standardized process in which cybersecurity and risk management activities can be integrated into the system developmental life cycle, encouraging consideration of cybersecurity where it is most effective – system design and development [8].

The RMF process consists of six steps shown in Figure 1. First, the system is categorized by identifying the information processed by the system as well as the impacts of the loss of this information. The criticality analysis of the system's information plays into the second step, in which security controls (or countermeasures) are selected. These

selected controls are implemented in step three. The fourth step then assesses how effective the security controls are through threat mapping (i.e., linking possible attack scenarios to weaknesses in control selection) and vulnerability analysis. Findings are compiled into a risk assessment and mitigation plan, which is briefed to an Authorization Official in step five. The final step of the RMF process is monitoring of the system with respect to implemented controls [8].

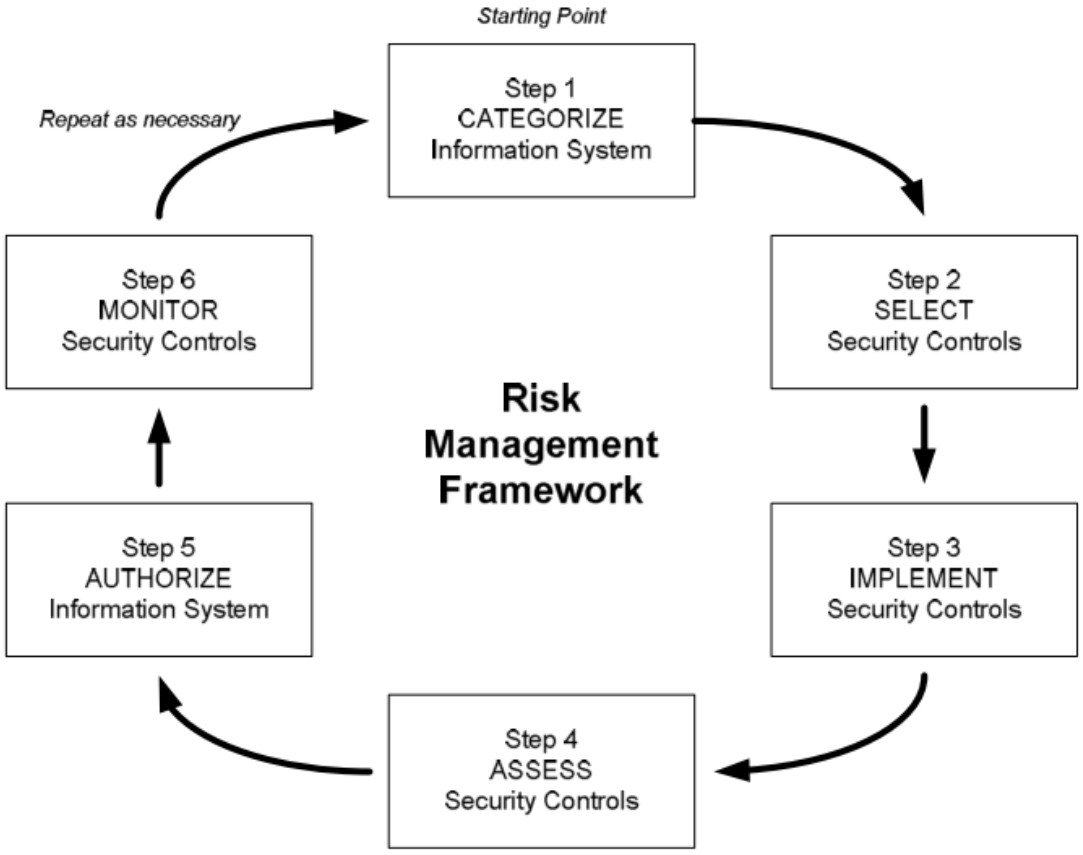


Figure 1. Risk Management Framework [8]

The Air Force Life Cycle Management Center (AFLCMC) has tailored the RMF to the systems for which they are responsible, primarily the operation and sustainment of legacy MWSs. This tailored process is called the Platform Information Technology Assessment and Authorization (PIT A&A) Process. This AFLCMC-unique process applies no modifications to the RMF steps, but rather shifts the focus. With legacy MWSs, the time for cybersecurity considerations during development has long since passed. For this reason, the PIT A&A requires the completion of steps 1-3 (Categorize, Select, and Implement) along with the resulting artifacts, thus emphasizing system Assessment and Authorization [9].

Unfortunately, this PIT A&A process is often accomplished with a compliance-based mindset instead of thinking critically about the unique threats and risks faced by individual MWSs [10]. Consequently, critical cybersecurity assessment and mitigation steps are often conducted simply to comply with policies and procedures instead of applying a risk-based approach to develop safe and secure systems. In addition, a fundamental problem occurs after systems have been authorized. Organizations should be employing continuous monitoring of the cybersecurity state of the system [8]. However, few look past the title of step 6 (Monitor Security Controls) and all that ends up being monitored are selected controls.

Another challenge organizations face in completing the PIT A&A process is developing expertise to perform analyses and detailed assessments of the various systems of interest (SoI) to provide the necessary evidences of thorough analysis and engineering rigor (e.g., artifacts and traceability). Because of difficulties and limitations in executing

the PIT A&A process, required rework imposed, and process duration, the Authorization Office has a backlog of over 250 non-compliant systems.

Problem Statement

Compliant weapon systems should not be the goal, but rather a byproduct of effective risk assessment and mitigation practices which produce secure systems. This is challenging when compliance with directives is telegraphed as the primary objective [10]. Because of the increasing concern that military systems and their underlying information systems are vulnerable to exploitation and offensive attack through cyberspace, it is essential that DoD organizations accomplish the assessment and authorization process with the goal of identifying and mitigating risks in an efficient and effective manner.

Research Questions

The following research questions focus on helping organizations and system owners understand the benefits of the PIT A&A Process. Additionally, it is a goal to help assist in shifting the DoD's emphasis from compliance to mission assurance and ultimately improve the state of MWS cybersecurity.

1. How can the PIT A&A Process be studied using modeling and simulation?
2. How can a baseline for measuring the effectiveness of the PIT A&A Process be established?
3. How does workforce proficiency, accuracy thresholds or security artifacts, and difficulty of artifact completion impact the PIT A&A Process's ability to identify and mitigate cybersecurity risks?

4. How can the PIT A&A Process maximize risk mitigation while minimizing artifact rework?

Methodology

Simulation and stochastic models allow analysts to test processes under multiple sets of model specifications (i.e., input parameters and/or structural assumptions) to optimize performance [11]. The PIT A&A Process is a prime candidate for this type of analysis. This work identifies critical factors that drive the PIT A&A Process's ability to identify and mitigate risk through the development and analysis of a mathematical simulation through stochastic modeling. With an understanding of key drivers in risk identification and mitigation, factor modification experiments were conducted to improve process effectiveness. This was accomplished through a design of experiments (DOE) to find factor level configuration that maximizes risk mitigation while minimizing the amount of rework required. Development of the PIT A&A stochastic model was accomplished in Microsoft Excel.

Preview

In order to further understand and improve the process of developing secure and defensible weapon systems, this work conducts a detailed study of AFLCMC's PIT A&A process and its ability to identify and mitigate cybersecurity system risks faced by DoD weapon systems. Chapter II provides a history of DoD's efforts to apply risk management to conventional and platform information systems as well as an in-depth description of the RMF process. In Chapter III, a description of the stochastic model is explained along with critical modeling assumptions. Chapter IV provides a detailed discussion of the

experiment results with the goal of maximizing risk mitigation while minimizing the amount of rework required. Chapter V offers conclusions and suggests future work to further understand and improve risk management efforts in the development and fielding of secure weapon systems.

II. Literature Review

Chapter Overview

Within this chapter, definitions of cybersecurity and their application risk management are established. A brief history of DoD efforts to implement risk management to platform systems through various processes is presented. Need for Platform Information Technology (PIT) Assessment and Authorization (A&A) Process for legacy aircraft is explained and policies surrounding the PIT A&A Process are described. In addition, an in-depth review of each step in the PIT A&A process is conducted with the NIST Special Publication 800-37 as the primary source document on how the process should be completed.

Cybersecurity

Cybersecurity has the attention of the nation's most senior leaders. However, despite its importance and frequency as a topic of discussion, there is still is a lack of clarity and apparent understanding of the term cybersecurity. In an effort to bridge this cybersecurity knowledge gap, P.W. Singer and Allan Friedman help clarify this often misunderstood term in their book "Cybersecurity and Cyberwar: What Everyone Needs Know" [12]. They suggest that cybersecurity is preventing adversaries from gaining something through malicious activity – be it accessing private information, undermining the system, or preventing its legitimate use. The common perception is that "security" suggests freedom from danger, however it is more associated with the presence of an adversary. Cybersecurity is not a state in which a system is completely immune to attack, but rather an assurance that appropriate protections and mitigations are in place. This

coincides with the National Security Presidential Directive on Cybersecurity Policy which defines cybersecurity as [13]:

> "the prevention of damage to, protection of, and restoration of computers, electronic communication systems, electronic communications services, wire communication, and electronic communication, including information contained therein, to ensure its availability, integrity, authentication, confidentiality, and non-repudiation".

This definition was so well received that it was implemented into the Department of Defense's official instruction on cybersecurity, DoDI 8500.01 [14].

Risk Management History

Behind definitions, there still lies a major concern that DoD systems containing cyber capabilities are vulnerable to corruption and attack, and this concern is on the rise. Because the DoD relies so heavily upon cyberspace to enable military operations, exploitation of vulnerabilities could undermine mission assurance and ultimately threaten national security [15]. While the cyber domain facilitates the use of innovative capabilities and offers a competitive edge over adversaries, connecting national and military infrastructures also provides access opportunities to practically anyone from any location around the world [16]. It is essential that cyber systems, particularly MWS, undergo some form of risk assessment and authorization to ensure cybersecurity through application of appropriate protections and mitigations.

The DoD recognized the need for such a process in the early 1990's. The Assistant Secretary of Defense for Command, Control, Computers, and Intelligence

issued the Defense Information Systems Security Program Strategic Plan, which created a standardized process for accrediting computers, systems and networks [17]. Public law was also written to ensure information security, specifically 44 United States Code Chapter 35, Subchapter III on information security, which states [18],

> "Each agency shall develop, document, and implement an agency wide information security program… to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source, that includes … periodic assessments of the risk and magnitude of the harm that could result from the unauthorized access, use, disclosure, disruption, modification, or destruction of information and information systems that support the operations and assets of the agency."

This mandate inspired the development of an approach to information system security certification and accreditation. The instruction came formally through DoD instruction 5200.40, Department of Defense Information Technology Security Certification and Accreditation Process (DITSCAP) [19]. This process provided a streamlined method in applying best practices, sound software design and proven security practices [7]. Despite this valiant attempt to improve the overall security posture of information systems, major inadequacies were revealed. One significant concern was that system interaction within the larger enterprise was not considered. Each system was treated as an independent silo. In addition, a standard list of controls had not been created [20]. In an attempt to improve this process, DITSCAP was officially replaced in 2007

with the Defense Information Assurance Certification and Accreditation Process (DIACAP).

DIACAP retained a focus on individual systems, but was much more enterprise-centric [21]. An enumerated standard control set was adopted (found in DoD 8500.2 [22]), and paperwork requirements were streamlined. Notwithstanding the improvements from DITSCAP, another blaring issue remained – the DoD was using a completely different authorization process and control sets than the rest of the Federal Government. Without some type of translation process, interconnectivity between these systems was all but impossible [20].

In 2014, the DoD resolved this problem along with many others by shifting the emphasis from information security under the DIACAP to the goal of cybersecurity using the Risk Management Framework (RMF) [10]. The RMF provides a standardized process and a common control set with which cybersecurity and risk management activities can be integrated into system developmental life cycles across all federal agencies. This requires system and program managers to consider "baked-in" security in the development phase of the system life cycle. RMF also provides a more standardized language and approach at measuring cybersecurity risk. It promotes reciprocity across agencies under a single process, thus avoiding time consuming and costly re-authorization [23]. The new DoD RMF pioneered the first explicit guidance to address Platform Information Technology (PIT) and PIT systems with respect to cybersecurity [10]. The Air Force Lifecycle Management Center (AFLCMC) has responded to this direction in its development of the Platform Information Technology (PIT) Assessment and Authorization (A&A) Process.

PIT Definition

To fully understand what is being assessed and authorized within this process, the Information Assurance (IA) PIT Guidebook defines a platform as a "vehicle, structure or person that performs a special purpose mission in support of United States National Security policy; and aboard or in which a DoD national security system may be installed to support assigned missions" [24]. For the purpose of this work, the focus will primarily be on Aircraft and more generally Major Weapon Systems (MWS).

The Guidebook further defines Platform Information Technology as "a special purpose information system which employs computing resources (i.e., hardware, firmware, and optionally software) that are either physically embedded in the platform; or the information system has a special-purpose mission dedicated to supporting real-time mission performance of the platform" [24]. PIT Systems can be categorized as government-designed systems that contain elements cyber as well as architectures, protocols, and interfaces over which the government has as some degree of control [10]. MWS fit this description, where Commercial Off the Shelf (COTS) or general-purpose systems used to conduct routine administration or business applications would not considered to be PIT.

The primary issue in securing the majority of the DoD's PIT Weapon Systems is that very few were developed and fielded in an era when the notion of "cybersecurity" was even conceived [10]. However, new and legacy systems increasingly rely on cyber capabilities to carry out mission-essential tasks and maintain uncompromised operation. Because these systems may not have cybersecurity protections "baked in" or "bolted on,"

they must be assessed and authorized for use. The PIT A&A Process is the method by which this is accomplished.

RMF Process Overview

The PIT A&A Process is a tailored version of the RMF. It still follows the standard 6 steps found in the framework beginning with categorization of the system. This is done by identifying the information processed by the system as well as the impacts of the loss of this information. This criticality analysis plays into the second step, in which security controls, or countermeasures to avoid or minimize risk, are selected. The third step is implementing these selected controls. Within the fourth step, assessment of the security controls' effectiveness is conducted through threat mapping and vulnerability analysis. Findings of this analysis are compiled into a risk assessment and mitigation plan, which is then briefed to an Authorization Official to achieve step five. The final step of the RMF process is continuous monitoring of the system with respect to cybersecurity [8].

Because the PIT A&A Process is centered around legacy MWS, the entry criteria for the process is steps one through three must already be complete. In addition, the resulting artifacts must be available before assessment and authorization is considered [9]. Figure 2 provides a depiction of flow of each step within the RMF process, and the tasks associated with each step. The primary six RMF steps are listed across the top and decomposed into numbered tasks in orange and green. Key decision points are highlighted with purple diamonds. Specific actions within tasks or after decision nodes are represented with unnumbered light green rectangles. While all steps are conducted in

the PIT A&A process, the primary focus is on steps 4 and 5 as depicted with the red rectangle.

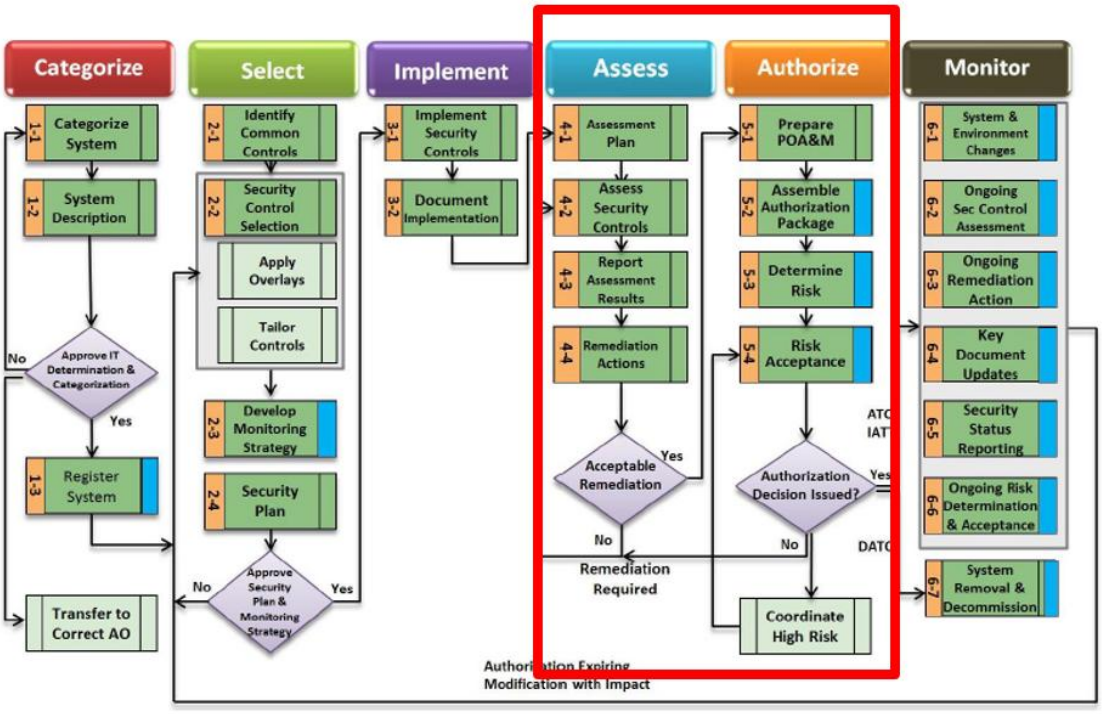


Figure 2. PIT A&A Process Task Flow [25]

Figure 3 identifies the artifacts or documentation produced within each step, however the PIT A&A specifically requires those artifacts identified with the red rectangles. Each step generates one or more artifacts to document the completion of activities that identify or mitigate risk. All artifacts are typically produced in the form of a multi-page Microsoft Word document. Supplemental data or figures may be analyzed or created using various software tools. For example, Department of Defense Architecture Framework (DoDAF) visualizations of system infrastructure (presented in the AAR) may be generated using specialized software (e.g., Cameo System Modeler). Results are then populated and presented in the Microsoft Word format.

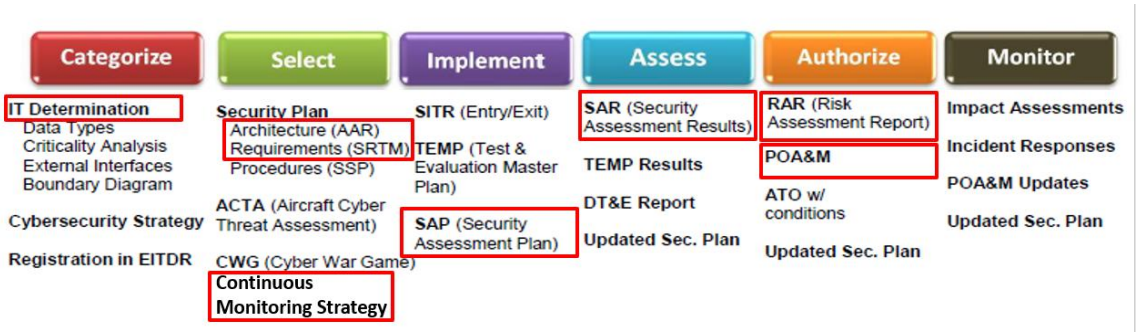


Figure 3. PIT A&A Process Deliverables [25]

To further understand these steps and tasks, the following sections thoroughly define how tasks are executed and what artifacts are produced as a result. The following sections are summaries of the 6 RMF steps as detailed in the NIST Special Publication 800-37: Guide for Applying the Risk Management Framework to Federal Information Systems [8].

- **Step One**

The first step is to categorize the system. This is an organization-wide activity initiated by the system owner and carried out in conjunction with appropriate organizational officers. The first task in step one is categorization and documentation, which is done by submitting an IT Determination package to the authorization official. Basic information that should be provided in the system description includes:

- Full descriptive name of the system along with its acronym
- Unique information system identifier/code
- Name and contact information for the information system owner
- Name of the system's governing organization
- Location of the system and environment
- Version number

Beyond this basic information, the description should/could include information such as:

- Purpose and capabilities of the system
- Type of information that is processed, stored, and transmitted via the system
- Status of the system in the development life cycle
- Boundaries for risk management purposes
- Encryption techniques used

For PIT systems, a PIT Determination checklist is provided. If the decision authority concurs they system is PIT, they accept responsibility for accrediting the system. Otherwise, the system is sent to a more appropriate authorization office [13]. Additional information may be included and/or required by the decision authority in the system description whenever it becomes available during the RMF life cycle.

Next, the system, including the system boundary, must be described. Descriptions of system unique characteristics such as data flows and types, external and internal interfaces, hardware and software inventory and any others are included so the system can be properly assessed. This is captured in the Architectural Analysis Report (AAR), which also contains the system's architecture and cybersecurity concerns [13]. This descriptive information should be included in security plan's system identification section or included in attachments. The amount and depth of information required in these descriptions is dependent on organization needs and correlates with the security categorization.

The final task in step one is to register the system in the Enterprise Information Technology Data Repository (EITDR). PIT systems are also required to be registered under this repository and follow the Warfighter Mission Area and PIT track in answering

required questions. This should be done by the system owner or program manager, who should request access to the EITDR and link the system with the appropriate program or management offices [13]. Registering the system establishes the relationship with the governing organization [8].

- **Step Two**

Step two in the process is to select security controls. The first task therein is to identify common controls for information systems as well as those provided by the organization, and then document these controls in a security plan. Common controls are inherited by one or more organizational information systems. For example, Common Access Card (CAC) is a typical control for network systems. There are over 900 controls prescribed to protect the confidentiality, integrity, and availability of the system. Below is a table of the various families of controls and their respective classes (see Table 1). Because of the number of controls, identifying common controls offers potential organizational cost savings in terms of level of effort [25].

IDENTIFIER	FAMILY	CLASS
AC	Access Control	Technical
AT	Awareness and Training	Operational
AU	Audit and Accountability	Technical
CA	Security Assessment and Authorization	Management
CM	Configuration Management	Operational
CP	Contingency Planning	Operational
IA	Identification and Authentication	Technical
IR	Incident Response	Operational
MA	Maintenance	Operational
MP	Media Protection	Operational
PE	Physical and Environmental Protection	Operational
PL	Planning	Management
PS	Personnel Security	Operational
RA	Risk Assessment	Management
SA	System and Services Acquisition	Management
SC	System and Communications Protection	Technical
SI	System and Information Integrity	Operational
PM	Program Management	Management

Table 1. Security Control Classes, Families, and Identities [26]

Unlike step one, in which the primary responsibility for tasks falls on the information systems owner, step two's responsibilities are on the shoulders of a person like the Air Force Chief Information Security Officer (CISO). The information system owner can now act in a supporting role along with others like the Information System Security Engineer (ISSE).

The next task is to select which controls to implement for the system and document them. The kinds of security controls required will be determined based on the system's security categorization. The selection process includes:

- Choosing a set of baseline controls
- Tailoring the baseline security controls by applying scoping, parameterization, and compensating control guidance
- Supplementing the tailored baseline controls with additional controls

- Specifying minimum assurance requirements

The security plan contains an overview of the security requirements and should provide enough detail to determine whether the selected controls will be sufficient to meet the system owner's security needs. In order to enable compliant implementation of the chosen controls, the security plan should describe the intended application of each control in detail specifically in the context of that system.

While the security controls are being selected, work may begin on the third task: developing a strategy for continuous monitoring of the security controls' effectiveness. Continuous monitoring, and the ability to make adjustments as needed, is crucial for maintaining cyber security. The most effective way to ensure an effective monitoring strategy is to develop and implement it early in the system development life cycle. A successful monitoring program should include:

- Configuration management and control processes
- Security impact analyses on proposed or actual changes to the system and its operational environment
- Assessment of selected security controls employed within and inherited by the system (including controls in dynamic subsystems)
- Security status reporting to appropriate organizational officials

The continuous monitoring strategy needs to spell out all the pertinent details of monitoring both the initial controls and any changes made to them. For the former, the strategy should identify the control, state how frequently it should be monitored, and how it should be assessed. These protocols and the frequency in which they are conducted will be determined by how important the information system is and how trustworthy the

controls (or control providers) are. For the latter, the strategy should describe how to monitor changes in the system and how to analyze the security impact of these changes. For both initial controls and changes made, the reporting requirements should detail what needs to be included on the reports, how often, and who should receive them.

The final task in step two is to review and approve the security plan. An independent review by authorization officials should determine whether risk was properly assessed and addressed by the security plan. If any issues are found, changes should be recommended and then implemented by the system owner. If potential risk is properly identified and the security plan is satisfactory, the plan should be approved [8].

- **Step Three**

Once the security plan is approved, the system can move to step three, which is to implement the security controls. This task should be thoroughly outlined in the approved plan. Security engineers are to use best practices when implementing the plan, to include software engineering methodologies, security engineering principles, and secure coding techniques. They should use a sound process that "captures and refines security requirements and ensures the integration of those requirements into information technology products and systems through purposeful security design or configuration" [8]. Mandatory configuration settings should be configured and implemented according to organizational and federal guidelines.

Next, the security control implementation should be documented in the security plan. This documentation formalizes expectations and should provide a functional description including planned inputs, expected behavior, and expected outputs. The

security plan should detail the information and security engineering methodologies used as well as which information technology products were integrated. Minimum assurance requirements should also be addressed to ensure compliance [8].

- **Step Four**

Now that the security controls have been implemented, step four is to develop, review, and approve a plan to assess the security controls. "The security assessment plan provides the objectives for the security control assessment, a detailed roadmap of how to conduct such an assessment, and assessment procedures" [8]. This should be done by the Security Control Assessor (SCA) who is appointed by the Air Force Chief Information Security Officer (CISO). The SCA is responsible for providing recommendations on risk assessments to identify any residual risk of the system [9]. This is done alongside the acquisition process to identify and correct weaknesses early in a cost-effective manner. This assessment identifies whether or not the controls were implemented correctly and to what extent. Controls should be operating as intended while producing the desired outcomes. The security control assessment plan provides the appropriate framework for this type of assessment to be conducted.

The first task in this step is to create the Security Assessment Plan, which establishes expectations and bounds the level of effort required – including expertise and independence – for the assessment. Once the plan is created, the next step is to assess the controls in accordance with the plan. They should have access to the system and environment of operation as well as all appropriate documentation – records, test results, etc. The assessors will be able to provide an unbiased, factual reporting of any

deficiencies found during their assessment. The independent assessors assess the security controls and provide specific recommendations on how to correct weaknesses or eliminate vulnerabilities (i.e., mitigations). When the independent assessment is complete, the SCA should oversee the preparation of a security assessment report that documents the issues, findings, and recommendations made by the assessors. Finally, initial remedial actions should be taken to improve the security controls based on the security assessment report [8].

- **Step Five**

Step five is to authorize the system. The first undertaking in this process is to prepare the last of three key documents (the first two being the security plan and the security assessment report) in the security authorization package – the plan of action and milestones (PO&AM). The PO&AM is based on the findings of the security assessment report (less the initial remedial actions already taken). It outlines the specific tasks required to correct weaknesses or deficiencies and address the system's residual vulnerabilities. The POA&M identifies four points:

- The task to be accomplished with a recommendation for completion either before or after system implementation
- The resources required to accomplish the tasks
- Any milestones in meeting the tasks
- The schedule completion dates for the milestones

If all identified issues were mitigated during the initial remediation, a plan of action and milestones is not required and the organization can go straight to the second task –

assembling the security authorization package. The security authorization package should include the security plan, the security assessment report, and the plan of action and milestones (if applicable). Authorizing officials use these three key documents to make risk-based authorization decisions. The authorizing official or their designated representative is responsible for determining the risk to organizational operations, assets, individuals, or the Nation. The explicit acceptance of risk, however, can only be made by the authorizing official. They must balance security considerations with operational needs, and if they deem the risk acceptable, the system can be authorized. The authorization decision document will convey the final security authorization decision.

- **Step Six**

Step six is to monitor implemented security controls where the system owner determines the security impact of proposed or actual changes to the system and its environment of operation. Of note, this step requires ongoing assessments of the security controls in accordance with the defined monitoring strategy (e.g., an annual assessment or revisiting of the ATO package). Remedial actions should be taken as the need arises based on the results of the ongoing monitoring. As the monitoring process continues, key updates need to be made to the security plan, security assessment report, and plan of action and milestones. Security status reports need to be provided to the authorizing official and others as previously outlined in the monitoring strategy (step two). Continuous monitoring requires the authorizing official to review reports and determine acceptable risk on an ongoing basis (e.g. annually). Lastly, when the system is

reaching its end of life, an information system decommissioning strategy may be implemented.

Summary

This chapter detailed the innerworkings of the PIT A&A Process to provide understanding of how each step should be executed. The DoD's past efforts to implement an information system/cybersecurity assessment and authorization process were reviewed to help gain a greater respect for the Risk Management Framework and its ability to be tailored to Major Weapon Systems with the PIT A&A Process. Cybersecurity definitions were well established and help set the stage for risk assessment expectations – creating safe and secure cyber systems with the goal of mission assurance. The next chapter lays out the methodology behind stochastic model development.

III. Methodology

Chapter Overview

This chapter introduces the 12 Steps in a Simulation Study which was used in developing a stochastic model of the PIT A&A Process. Each step is detailed along with a thorough explanation of how the model was created. This includes an in-depth description of model inputs, outputs, factors and calculations in addition to critical assumptions. Verification and validation exercises are explained to establish confidence in model accuracy, and model experiment designs are described.

Model Development

The innerworkings of a process responsible for ensuring appropriate measures are in place to protect MWSs from cybersecurity threats are complex. In order to understand how effective and efficient this process is at identifying and mitigating risks, major simplifications and assumptions about system behavior must be made. The results of the PIT A&A are very subjective, but in order to measure effectiveness and/or efficiency, process outcomes must be quantifiable. Because of the nature of the problem, the method by which this process should be studied lends itself to modeling and simulation. A model *is* a simplification of a real process and facilitates understanding, prediction and potentially control of process outcomes [27]. The knowledge gained in model development and simulation study can be of great value toward suggesting process improvements.

The next question is *how does one create a model in a simulation study*? Fortunately, in his book "*Discrete-Event System Simulation,"* Jerry Banks describes a

twelve-step iterative process that guides model construction [27]. These steps were used in creating the simulation model for the PIT A&A Process and will help guide discussion on how each step was accomplished. The 12 steps and their process flow can be seen in Figure 4 below.

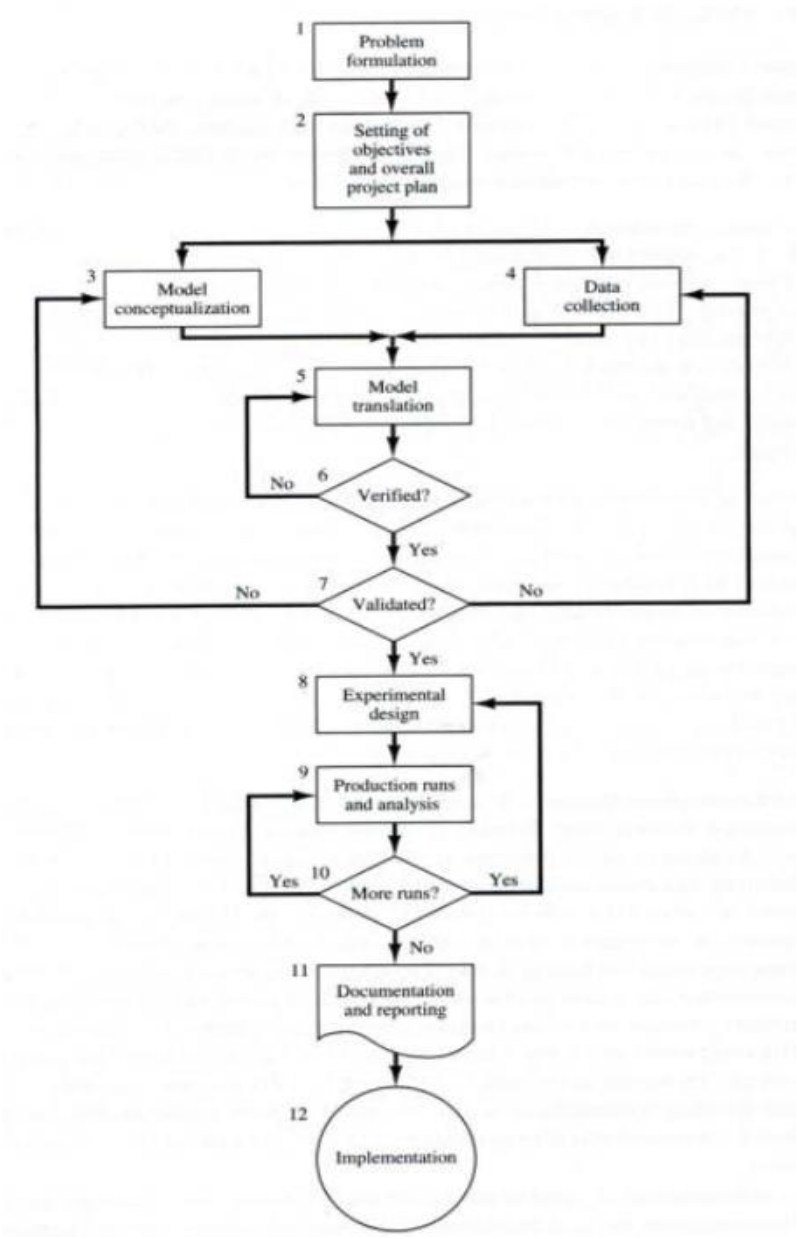


Figure 4. Steps in a Simulation Study [27]

Step 1 – Problem Formulation

The first of the twelve steps is problem formulation [27]. As previously established within Chapter I, the importance that MWSs be assessed with the goal of efficiently and effectively identifying and mitigating cybersecurity risks is ever increasing. This is due to the growing reliance upon cyber capabilities, the rising presence of maturing threats in cyberspace, and the severity of the impact if vulnerabilities are exploited.

Step 2 – Setting of Objectives and Overall Project Plan

A well-defined problem sets the stage for the second step - setting objectives. The objectives specify what questions will be answered by the simulation study [27]. Within the NIST 800-30, risk management starts with identification [28]. The goal of the PIT A&A process is to identify system risks, assess those risks, and take appropriate measures to reduce, mitigate, and ultimately eliminate those risks. It is assumed that each system has an unknown amount of risk that cannot be mitigated unless it has first been identified. The overall objective of the model can be seen in Figure 5. Given the risk associated with a system under assessment, the model should express the PIT A&A Process's ability to identify risk as a percentage of the total System Risk (as depicted in yellow). In like manner, the model should indicate how much risk is mitigated expressed as a percentage of the Risk Identified (shown in green). The goal of a PIT A&A Simulation Study is to understand and focus the PIT A&A Process on activities which help identify and mitigate cybersecurity risks in MWSs.

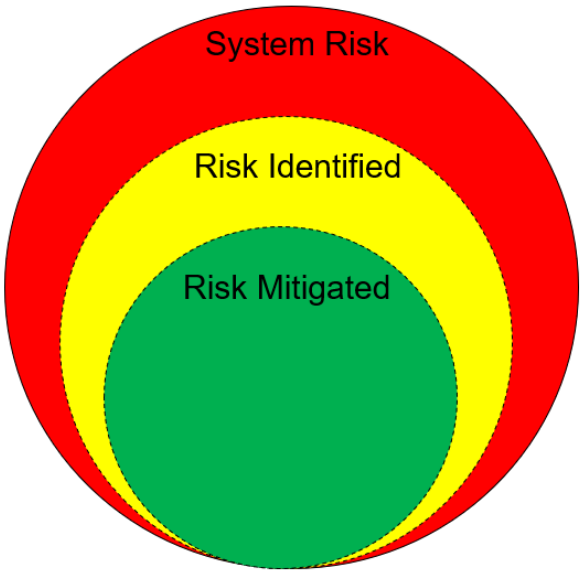


Figure 5. Model Concept - Risk Identification and Mitigation

Step 3 – Model Conceptualization

This design oriented step is considered to be as much an art as science. The challenge within modeling is capturing the essential features of the problem and selecting appropriate assumptions that appropriately characterize system behaviors of interest. Because of this, it is best to start with simple assumptions and a basic model, and add greater detail as needed in order to answer research questions. Complexity should not exceed what is required to accomplish the intended purposes of the model [27].

Assumptions/Limitations

As was mentioned in Step 2 – Setting Objectives, it is assumed that each system under assessment has an unknown amount of risk associated with it. It is also assumed that the amount of risk identified can be expressed as a percentage of the total system risk. The model only accounts for the risk identified within risk identification activities and their resulting artifacts, it does not account for any addition risk that may be

uncovered while conducting risk mitigation activities. In addition, it assumes the amount of risk mitigated can be expressed as a percentage of the risk identified. Although risk identification and mitigation are presented as percentages, they are not meant to be precisely interpreted as such; they are merely a means for measure and comparing for improvements.

Typically, the PIT A&A Process steps are executed in a fairly linear fashion over the course of many months. Because of this the artifacts were modeled independently with no cumulative effect in the amount of risk identified and mitigated. Note, accuracy of artifacts, employee comprehension, or personnel turnover could improve/diminish in later steps within the process. Therefore, no attempts have been made to model these effects. While the model may be restricted in reflecting outputs representative of the actual process timeline, the objective of this study is to understand which activities drive risk identification and mitigation and not to determine how long it takes to accomplish them. Modeling the artifacts independently permits better understanding of which artifacts directly impact risk identification and mitigation, even if process interactions are not represented in the model.

In order for a system to be authorized for use, the eight artifacts included within an authorization package must show sufficient accuracy in identifying system risks and ensuring appropriate measures are in place to mitigate or eliminate those risks. With respect to risk identification, the IT Determination and AAR generally provide detailed visibility into the system's overall risk posture. With respect to risk mitigation, the remaining six artifacts (SRTM, SAP, SAR, Continuous Monitoring Plan, RAR, and POA&M) provide insight as to how the risks can be mitigated. Note these artifacts do not

specifically mitigate the risks, but indicate how risks might be mitigated through successful completion of the PIT A&A Process, especially adherence to the POA&M through continuous monitoring. Because of this, the first and overarching assumption with the model is that accurate artifacts and their implementation equates to effective risk mitigation.

Currently, the PIT A&A Process does not include formal criteria to measure the accuracy of a given artifact. In this study, the Artifact Accuracy is intended to indicate both the appropriate level of detail and comprehensiveness of the artifact with respect to the system of interest. Thus, the model is limited in that it does not provide a precise artifact accuracy evaluation, but provides the necessary detail in order to meet the objectives of this modeling and simulation study.

For the purpose of this study, Artifact Accuracy is based upon three factors – System Complexity, Worker Proficiency, and Artifact Difficulty. Undoubtedly, there are many other factors that could impact the initial accuracy of each artifact, but it was desirable to keep model complexity low such that results can be more easily interpretable. Regarding System Complexity, it was assumed the more complex the system, the more challenging it is to achieve adequate artifact accuracy. A similar assumption was made with respect to Artifact Difficulty – the more difficult an activity and its resulting artifact, the harder it is to reach satisfactory accuracy levels. System Complexity and Artifact Difficulty are assumed to have a linear impact on artifact accuracy. Worker Proficiency levels are determined by years of experience as defined by AFLCMC [29]. Worker Proficiency levels also assumed a linear relationship with initial Artifact Accuracy (i.e., high proficiency yields a higher initial accuracy). This is because there is no historical

data documenting this increase, and thus, no way to determine how to measure it. These assumptions with respect to System Complexity, Artifact Difficulty, and Worker Proficiency do not negatively impact research objectives. For example, Worker Proficiency may increase over time, but it would not significantly alter the results of the simulation with respect to effectiveness (not impacted) or timeliness (slightly impacted).

The artifacts produced are rarely satisfactory after their initial draft. Multiple iterations, or reworks, are often required for most artifacts until they are deemed acceptable by the authorization office. It is assumed that each round of rework accomplished on a given artifact improves its respective accuracy. There is no data to support how frequently artifacts require rework nor to determine the improvement per rework; however, these details are not necessary to study which activities drive risk mitigation within the PIT A&A Process. For this reason, the improvement shown in artifact accuracy with each completed round of rework is assumed to be 10%. This allows artifact reworks to be studied in a general sense, where the model is not currently not capable of precisely predicting how long the PIT A&A Process takes.

Risk Identification/Mitigation

When working through the steps of the PIT A&A Process the progression of artifact development lends itself to the approach of risk identification first and then risk mitigation. In order to quantify how effective this process is at mitigating risk, there first needs to be a measure of how much risk can be identified. To do this, artifacts are categorized as either supporting risk identification or risk mitigation. This can be seen in

Figure 6.

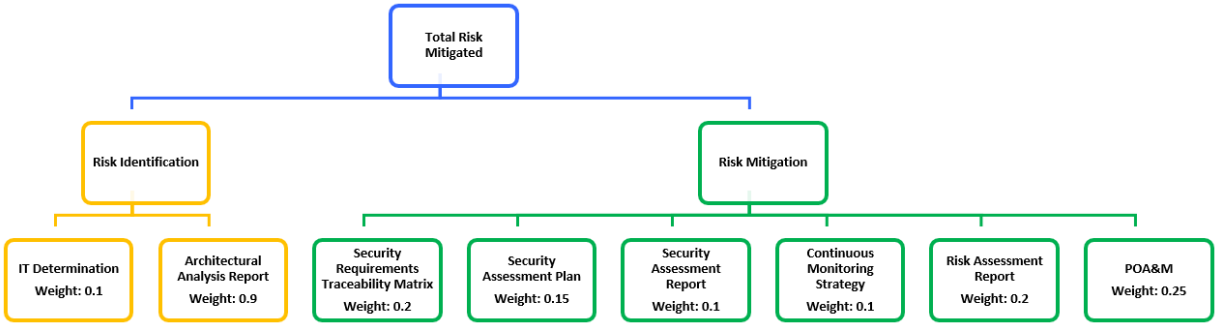


Figure 6. Artifact Identification/Mitigation Categorization

While Artifact Accuracy is the fundamental output of the process, the primary outputs of interest are Risk Identification and Risk Mitigation. These two primary outputs are dependent upon the sum of how accurately the artifacts are completed. However, not all artifacts are created equal. To show the importance of an artifact with either the identification of risk or the mitigation of risk, weights were assigned to each artifact based on their perceived level of importance. The weights of all risk identification artifacts summed to one, and likewise weights of all risk mitigation artifacts summed to one (values can be seen in Figure 6 and in Table 2). This allows Risk Identification to be expressed as a percentage of the total system risk. In like manner, Risk Mitigation can be expressed as a percentage of identified risk. To calculate how much risk can be identified, the calculated accuracy of each risk identification artifact was multiplied by their respective weights and added together.

$$RiskID = \sum(ArtAcc_{ID} * Weight_{ID}) \quad (1)$$

This number corresponds to the percentage of system risk identified, or the yellow circle depicted in Figure 5, of the total system risk.

In order to calculate how much of the identified risk can be mitigated, calculation of accuracy of each risk mitigation artifact is multiplied by their respective weights and summed together.

$$RiskMit = \sum(ArtAcc_{Mit} * Weight_{Mit}) \quad (2)$$

This value is the percentage of identified risk that was mitigated, or the green circle in Figure 5.

Total Risk Mitigated

In order to fully understand how much risk is being mitigated compared to the overall system risk, the last and main output of interest is total risk mitigated. This is simply calculated by multiplying the risk identification value with risk mitigation (see Equation 3 below).

$$TotalRiskMit = RiskID * RiskMit \quad (3)$$

This factor ultimately indicates how effective the process is at mitigating risk. The primary objective of this study is to maximize total system risk mitigated. This is done by maximizing the amount of risk identified and mitigating identified risk.

Before any computer code is written, the process flow must be devised and conceptualized [27]. This was done by walking through each step of the PIT A&A Process, identifying all inputs and outputs of the respective steps, understanding how these inputs and outputs and other factors affect other subsequent steps within the process. An example of an input-output diagram for a step within the process is

illustrated in Figure 7. Once the process flow has been conceptualized, a simple model can be created and built toward greater complexity as model refinement continues.

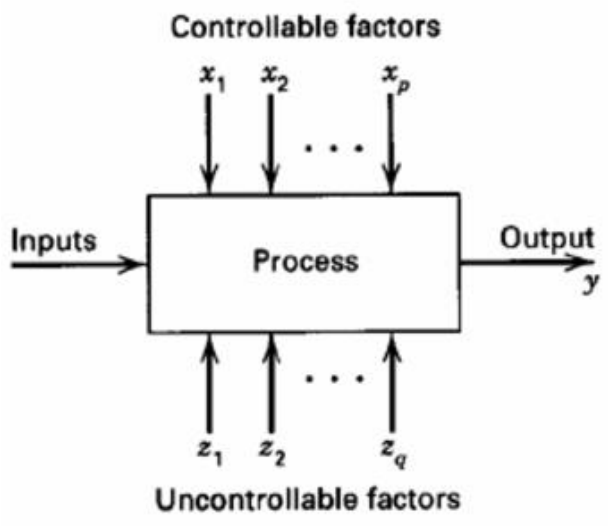


Figure 7. Input-Output Diagram [27]

System Complexity

The primary input to the PIT A&A Process is a Major Weapon System (MWS). There are a number of uncontrollable factors with respect to the assessment of MWSs – the first being System Complexity. The more complex the system, the more challenging it is to identify and understand the vulnerabilities and how best to mitigate risks associated with those vulnerabilities. For example, there is a substantial difference in the complexity of an F-22 compared to that of an AIM-9 Missile. Yet both of these "systems" are considered Platform Information Technology (PIT) and require assessment through the same steps of the PIT A&A Process. Although this is an uncontrollable factor, values for this metric need to be quantified respective to their impact on system outcomes.

For the purpose of this study, three levels of complexity are considered – high, medium and low. A diminishing effect on system outcomes was assumed as system complexity increased. The following values were deemed appropriate: 50% for high complexity; 75% for medium; and 90% for low. Because the authorization office does not have control over the type of system in need of assessment, this factor was modeled to be a random draw between the three complexity values. Within Microsoft Excel, the RAND function assumes a uniform distribution between the identified values.

Worker Proficiency

In reality, the next factor is very challenging to control; however, for the purpose of this study it is classified as controllable. The proficiency of those completing the process (or Worker Proficiency) has a direct effect on system outputs. Untrained and incompetent personnel could severely degrade the comprehensiveness of the system assessment. A more experienced workforce would have a better understanding not only of the systems themselves but also the meticulousness required to yield a satisfactory product. AFLCMC has identified 5 proficiency levels to reflect the various stages of worker competency [29].

Expert is the highest level and defines those who typically have 9+ years of wide breadth experience, anticipate and solve problems using mature judgement, and demonstrate high competency in exceptionally difficult situations. Workers with 6-8 years of experience are considered to be Advanced or Level 4. These are those who independently apply standard practices to perform complex tasks. Level 3 or Intermediate workers have 3-5 years of experience and require only occasional guidance, whereas frequent guidance is required for Level 2 or Basic personnel who characteristically have

6 months to 2 years of experience. This leaves Level 1 or Awareness employees who are essentially active trainees with 0 to 6 months of training. Level 1 workers still have some familiarity in the specific area, but their understanding is limited due to inexperience.

For this study, it was assumed that any employee regardless of their proficiency level could be tasked to populate the required artifacts because there is no requirement specifying worker competency. Similar to system complexity, values are assigned to each proficiency level to demonstrate the workers ability to successfully complete tasks. The following approximations reflect an interpretation of 5 proficiency levels and the impact values considered to be most fitting. Level 5 was assigned a value of 99%, level 4 – 90%, level 3 – 80%, level 2 – 70%, and level 1 – 60%. There have been no prior efforts made to quantify worker proficiency in terms of percentage.

Initially, the thought was to multiply each level by 20% (i.e., Level 5 – 100%...Level 1 – 20%), however, a Basic employee (Level 1) was thought to be more capable than submitting a 20% solution. In addition, it did not seem reasonable for Experts (Level 5) to achieve an initial solution of 100%. For these reasons, high and low-level percentages of 99% and 60% were selected with interim proficiency levels evenly distributed between these values. It was thought that approximately 10% between each level would show enough change in process results. Similar to system complexity, authorization offices have limited amount of control over who accomplishes the required documentation and their respective competency. To demonstrate this, worker proficiency was modeled to be a random selection of the 5 proficiency levels (again assuming a uniform distribution).

Artifact Accuracy

Before identifying the next controllable factors, it would be beneficial to first discuss process the fundamental output effected by these factors. Throughout the execution of PIT A&A Process, there are eight primary artifacts produced which make up an authorization package. The success of the authorization package's ability to mitigate risk is determined based upon how well the individual artifacts are constructed. Therefore, the model results in terms of risk mitigation are assumed to be a function of Artifact Accuracy. The accuracy of an artifact is expressed as a percentage or a score out of 100. The higher the percentage, the higher the quality of the document. Although artifact accuracy can be seen as very subjective, there are a number of factors that can have a real effect on product quality. The three primary factors that impact artifact accuracy are System Complexity, Worker Proficiency, and Artifact Difficulty. These impacts will be explored in Chapter IV.

Artifact Difficulty

Initially, Artifact Accuracy was determined by simply multiplying System Complexity and Worker Proficiency; however, another variable needed to be introduced to capture the amount of rigor and system understanding required to complete various artifacts. Each artifact was examined and given a difficulty rating on a scale of 0.10 to 0.50 based on the perceived effort required to populate that specific artifact. The higher the difficulty, the higher the rating. The final equation for artifact accuracy is shown in Equation (4). Subtracting the rating on this scale from one was used to avoid confusion of an inverse relationship between artifact difficulty and its rating (i.e., as difficulty goes up, the rating goes down).

$$ArtAcc = SysComp * WkrProf * (1 - ArtDiff) \quad (41)$$

Artifact Threshold

Equation (4) only determines initial accuracy of a given artifact. In reality, the majority of artifacts require revisions and undergo rework. For the purpose of this study, it was assumed that each time an artifact underwent rework the accuracy improved by 10%. It is also assumed artifacts go through multiple rounds of revisions until they met or exceeded its respective threshold. This introduces the last controllable factor – Artifact Threshold. Each artifact type is considered individually and assigned a threshold value. These values are allotted based on what was thought to be an appropriate level of fidelity to meet the intent of that specific artifact. The values selected for both Artifact Difficulty and Threshold can be seen in Table 2. These baseline factor levels were determined as a result of in depth study of the RMF and PIT A&A Process as depicted in the NIST Special Publication 800-37 [8], the PIT A&A Guidebook [13], and other supporting documentation.

Reworks

With initial artifact accuracy calculated and thresholds assigned, the number of rounds of reworks required can be determined for each artifact. This is done by taking the initial artifact accuracy, subtracting it from the threshold, and dividing by 10%.

$$\#Rwks = \frac{Thres - ArtAcc_{Initial}}{0.1} \quad (5)$$

This calculation introduces a principal output of interest. While this model focuses on the PIT A&A process's ability to identify and mitigate risk, it is also important to

consider the amount of work required to accomplish satisfactory risk identification and mitigation.

Step 4 – Data Collection

Step four is Data Collection, though there is a constant interplay between model construction and collection of input data [27]. In fact, data collection throughout the model development process helps verify the model is designed correctly. The data needed for the PIT A&A Simulation Study is essentially historical information indicating duration of assessments authorization approval for various systems. With these historical data, the model could validate how accurately (or inaccurately) it represents the actual process time. However, this would not validate risk mitigation results. Because no quantifiable historic data for risk mitigation is available, all data collection occurred after model completion.

Step 5 – Model Translation

Step five, Model Translation, is where the modeler enters the model concept into a computer program in order to run simulations [27]. For this work, the program used to construct the model and complete simulation runs on the PIT A&A Process was Microsoft Excel. Model construction began by selecting one of the three specified values for System Complexity. This selection was random using the RAND function within Excel. Similarly, a random selection between levels of Worker Proficiency was coded. Each artifact type was separated in different colored columns with their respective Threshold, Weight, and Difficulty values populated. Using these values and the randomly generated values of System Complexity and Worker Proficiency, the initial Artifact

Accuracies of all eight artifacts were calculated (see Equation 4). The initial values for Risk Identified, Risk Mitigated, and Total Risk Mitigated were then calculated in separate columns (see Equations 1-3). Next, the number of reworks required to meet Artifact Thresholds was calculated for each artifact (see Equation 5). If initial Accuracy of an artifact did not meet the respective Threshold, that artifact underwent a round of rework where 10% was added to the Accuracy percentage. Accuracy improvements were tracked with each rework round in separate columns and colored coordinated with the respective artifacts. In like manner, improvements in Risk Identification, Mitigation, and Total Risk Mitigated were calculated with each rework round. Multiple rounds were conducted until all Thresholds were met or exceeded.

Step 6 - Verification

Verification is step six and is concerned with building the model correctly and whether or not the model is performing properly. For the most part, common sense is used in completing this step; however, there are a number of common-sense activities that can be used in the verification process. First, have the model reviewed by someone other than its developer, ideally an expert in the simulation software [27]. Throughout the model development process, verification that model parameters and model structure were reasonable was conducted by multiple outside parties skilled in the simulation software.

Another activity with more tangible verification evidence is to closely examine the reasonableness of model output under a variety of settings of the input parameters. This was done by looking at the total number of required reworks while changing the level of worker proficiency input. Figure 8 demonstrates how the number of reworks

(represented by the blue line) decreases as worker proficiency increases. This means there is less rework required with a more competent workforce. A highly proficient individual should yield a higher quality product or more accurate artifact, thus requiring less rework to achieve an acceptable accuracy threshold. The increase in artifact accuracy at higher levels of proficiency is seen with the red bars. These represent the average initial accuracy of all eight artifacts achieved by the respective workforce proficiency level. The results of these verification activities indicate the model is working correctly.

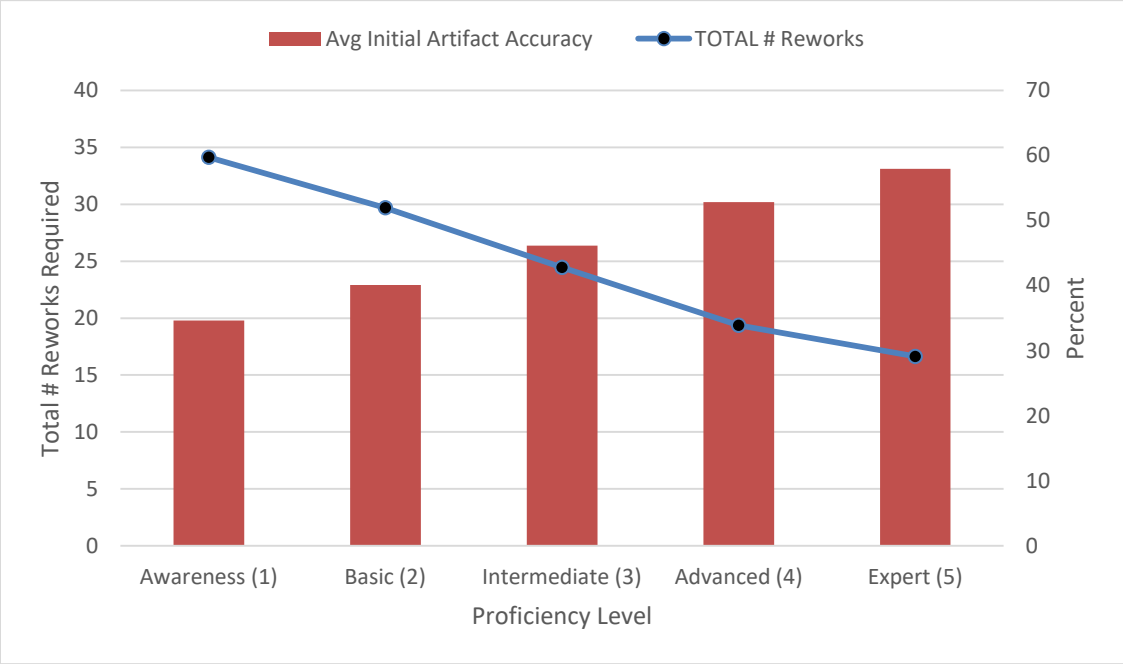


Figure 8. Verification of Reworks & Accuracy vs Proficiency Level

Step 7 - Validation

With Verification, one is confirming the model has been designed correctly, whereas Validation, which is step seven, is a determination that the model is an accurate representation of the actual process. No model is ever completely representative of the system or process under study; however, validation and calibration activities attempt to

bridge the gap between results from the model and the actual process. Within the validation step, a comparison of model performance is weighed against the actual process. Comparison tests are accomplished one of two ways – subjectively or objectively [27].

Subjective comparison tests involve people who are considered process subject matter experts, making judgments about the model and its outputs to ensure the simulation model is functioning properly and replicates the process accurately. After a conceptual model of the PIT A&A Process was presented, evaluations made specifically by process owners revealed the model's focus was centered around the wrong metric – process time. This resulted in a major shift in model development to examine effectiveness in risk identification and mitigation.

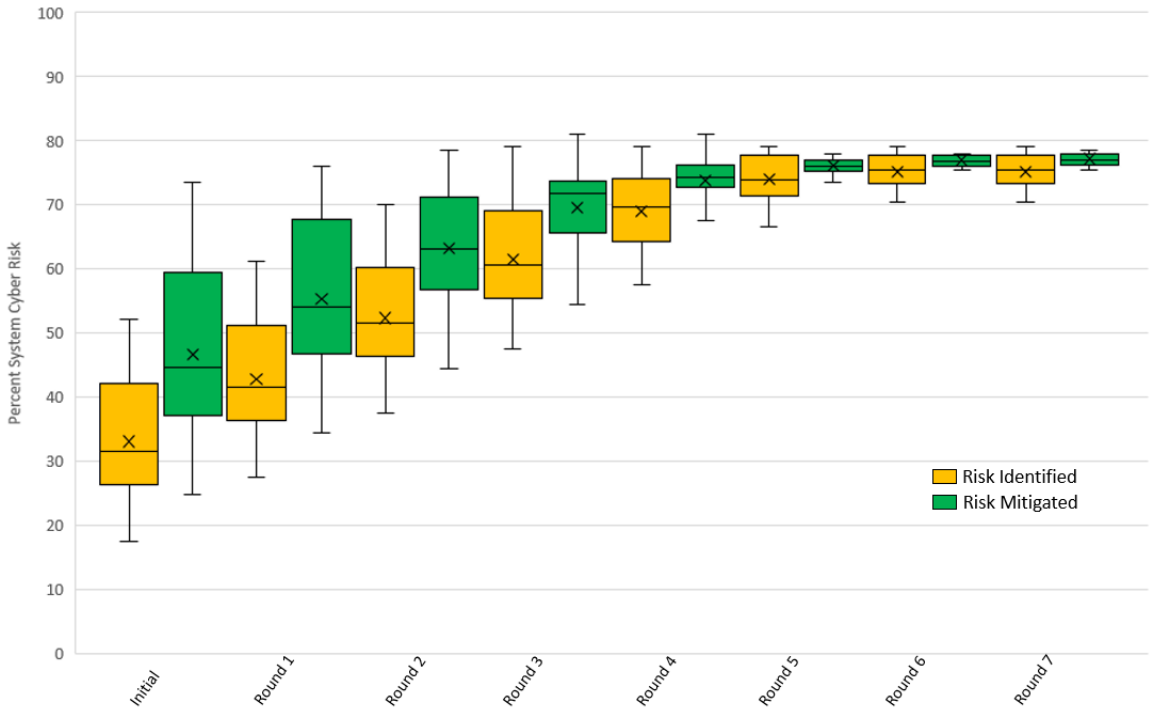
Objective tests require data from the actual process to compare to the model outputs. However, by nature, the PIT A&A process does not quantify its ability to identify and mitigate risk. Although, objective comparison tests cannot be conducted, a number of validation techniques can be executed to confirm the correct system was modeled.

One validation technique utilized was the use of operational graphics. Within this technique, values of various performance measures are depicted graphically during model simulation. This is done so model behaviors can be visually displayed to ensure proper performance [30]. An expected behavior of the PIT A&A model is the improvement of artifact accuracy, and thus risk identification and risk mitigation, as the artifacts undergo multiple rounds of reworks. This behavior can be visualized in Figure 9. The yellow boxes represent the range of risk mitigated and the green boxes equal the range of risk

mitigation over 1000 trials. There are initial values of risk identified and mitigated and with subsequent rounds of rework these values should increase. The average, maximum, and minimum number of reworks conducted each round is presented below the graph. As the number of reworks per round decreases, the improvement in risk identification and mitigation values should be less. This can be seen after Round 3 – the number of reworks per round drops off and risk identification and mitigation taper. These same interactions can be seen in Figure 10 where Total Risk Mitigated is considered with the blue boxes.

Another confirmation the model is working properly is illustrated by the fact that risk identification and total risk mitigated never achieve 100 percent, but converge on values between 70-80 percent. The model is designed in a way that limits artifact accuracy from improving beyond their designated threshold. Because the average threshold across all artifacts is approximately 73 percent, this trend in the data was expected.

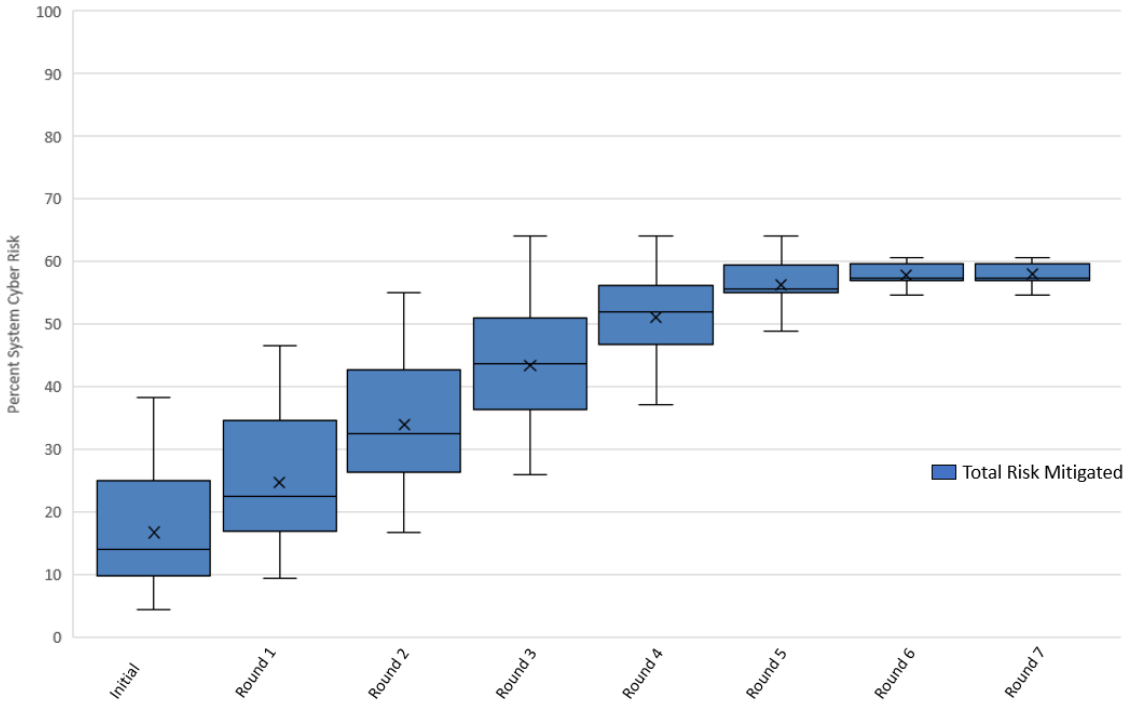
Another validation test that can be applied using these same figures is an internal validity test. Within this test, several replications of the model are conducted to determine the amount of variability in the model. Large variability identifies the lack of consistency and puts model results into question [30]. Figures 9 and 10 are box and whisker plots where the yellow, green and blue boxes represent the variance in Risk Identification, Risk Mitigation, and Total Risk Mitigated respectively over 1000 replications. As rounds of rework are accomplished, the boxes get smaller, i.e., variance decreases. With little variability resulting after all rounds of rework are complete, conclusions can be made that results are consistent and the model is valid.



 # Reworks per Round

Max	8	8	8	8	6	3	2	0
Avg	8	8	6	3	2	2	0	0
Min	2	2	2	1	0	0	0	0

Figure 9. Validation Risk Identification/Mitigation vs Reworks



# Reworks per Round								
Max	8	8	8	8	6	3	2	0
Avg	8	8	6	3	2	2	0	0
Min	2	2	2	1	0	0	0	0

Figure 10. Validation Total Risk Mitigated vs Reworks

Step 8 – Experimental Design

Once Verification and Validation of the model are complete, step eight is to conduct Experimental Design. Within this step alternative levels of factors of interest are simulated [27]. The controllable factors within the PIT A&A Process are Worker Proficiency, Artifact Difficulty and Artifact Threshold. Because Difficulty and Threshold apply to all eight primary documents with the PIT A&A Process, there would be an exponentially large number of factor level interactions to test. For this purpose, the two

most heavily weighted artifacts (Architecture Analysis Report and POA&M) and their respective thresholds and difficulties were included within the study.

To begin, baseline results will be collected with the factors of interest set at levels perceived to be representative of the actual process (see Table 2). These baseline results are found at the beginning of Chapter IV within Figure 11. Next, a sensitivity analysis will be conducted by modifying one factor at a time while comparing results to the baseline. Sensitivity analysis results can be reviewed in Chapter IV (see Figures 12-16). This will help determine which factors are the primary drivers in identification and mitigation effectiveness. Next, using these identified driving factors, multiple two-level factorial design of experiments will be used to show critical interactions between factors. Three different designs along with their results and interpretations are presented in Chapter IV (see Tables 3-8). Based off these results, four factor level configurations will then be selected for further study as potential process improvements. These configuration options are given in Table 9 and comparison results are presented at the end of Chapter IV (see Figures 17 and 18).

Artifact	Difficulty	Threshold	Weight
IT Determination	0.1	0.6	0.1
Architecture Analysis	0.45	0.7	0.9
Security Requirements Traceability Matrix	0.1	0.7	0.2
Security Assessment Plan	0.15	0.7	0.15
Security Assessment Report	0.2	0.7	0.1
Continuous Monitoring Strategy	0.05	0.6	0.1
Risk Assessment Report	0.1	0.7	0.2
POA&M	0.35	0.7	0.25

Table 2. Artifact Baseline Values

Step 9 – Execute Production Runs and Analysis

Now that the experiments have been designed, step nine is to execute the production runs and conduct analysis on the results [27]. Analysis of results is discussed in detail in Chapter IV.

Step 10 – More Runs

Given the analysis of completed runs, step ten is then determining whether or not additional runs are needed. If the analyst sees fit to conduct more simulation runs, it is also determined which design configuration those experiments should follow [27]. Through the Validation and Verification of the model, it was determined 1000 replications or runs was more than sufficient to complete appropriate analysis of the model.

Step 11 – Documentation and Reporting

Step eleven is actually two-fold: program documentation is recording how the simulation computer program operates in order to provide some form of continuity; and progress documentation provides a chronology of work done, decisions made, and the overall progress of the simulation model. On the reporting side, frequent deliverables accounting for milestones and accomplishments are submitted over the life of the project [27].

Step 12 - Implementation

Lastly, implementation of the recommended changes based on analysis is conducted. The success of this step is largely dependent upon the owning organization and is outside the scope of this effort [27].

Summary

Ultimately, following the aforementioned steps in creating a simulation model of the PIT A&A Process should yield inefficiencies as well as recommended changes. These changes will play an important role in not only creating more secure weapons systems but also ensuring existing systems are more cyber resilient. Understanding the use of modeling and simulation in system process analysis is a powerful tool. Knowing when to apply this tool is also critical. Because the of the complex decisions throughout the PIT A&A Process, as well as the duration of one process cycle, modeling and simulation is a very useful tool in discovering where the process can be improved. The assistance of the twelve steps in simulation model creation provide a sound methodology in creating a process model for the PIT A&A Process. Within this chapter, these twelve steps were described and thorough explanation of how the PIT A&A Process was modeled to measure risk identification and risk mitigation was provided.

IV. Analysis and Results

Chapter Overview

First, this chapter presents baseline model results representative of the PIT A&A Process. Next, sensitivity analysis is conducted by modifying individual factor levels within of the model to determine critical factors and interactions. With understanding of basic model interactions, a select number of two-level factorial configurations are further studied for potential process improvements. Lastly, based on the interactions within the different experiment designs, 4 configuration options are presented. Statistical analysis of each option in comparison to baseline results are presented and analyzed.

Baseline Results

Within the previous chapter, baseline values for process factors were established to provide a representation of the current state of PIT A&A Process (see Table 2). After simulation is complete at these factor levels, baseline results indicate how well the current process identifies and mitigates risk. It should be noted that these results are constrained by the assumptions made within the simulation model (see chapter III for details). In reference to the "Model Concept - Risk Identification and Mitigation" (see Figure 6), Figure 11 provides a good visual of how much of the total system risk is identified and mitigated. The red portion classifies the total system risk, the yellow section distinguishes how much risk is identified with the green representing the total amount of risk mitigated based on the modeled PIT A&A Process. A thousand replications of the entire PIT A&A Process model were simulated and the average of these primary outputs are displayed. This radar plot progresses through an entire cycle of

the process illustrating how the amount of risk identified and mitigated improves with each round of reworks. Seven rework rounds were conducted for each replication; however, once an artifact's accuracy met or exceeded its respective threshold, no additional rework is conducted on that particular artifact. The decision to conduct seven rounds was based on the maximum number of reworks required to meet all artifact thresholds (shown in Table 2).

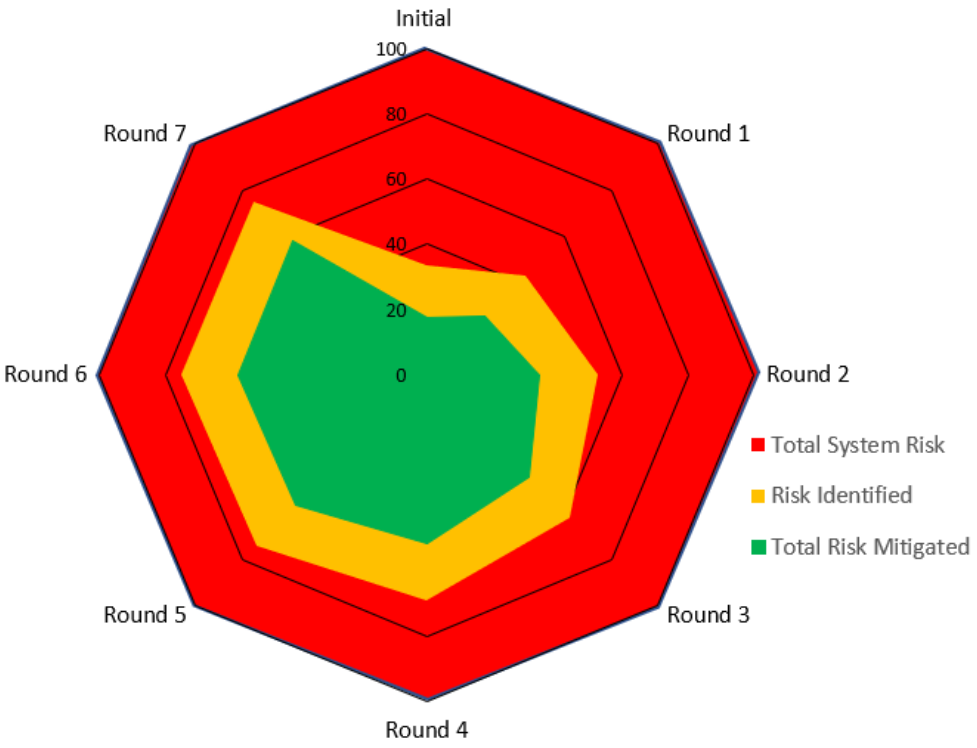


Figure 11. Baseline Radar Risk Identified and Total Risk Mitigated

Figure 11 shows that based on modeling assumptions, the PIT A&A Process can effectively identify approximately 75% of the total system risk. Of the risk identified, the process can successfully mitigate 77%, meaning roughly 57% of the total system risk is mitigated shown in green. The remaining 43% of the total system risk unmitigated seems

fairly poor at first glance. However, considering over half of this is assumed to be unidentified, 20% risk remaining is reasonably satisfactory. It is challenging to determine what is sufficient in terms of cybersecurity. In his article *"Good Enough" Security: The Best We'll Ever Have*, George Hurlburt suggests a completely secure system would most likely not be connected to the internet preventing any type of malicious activity from outside sources [31]. Essentially, a "completely secure system" would serve absolutely no useful purpose.

In reality, the amount of unmitigated risk deemed acceptable should be determined on an individual system basis when compared to the mission assurance requirements associated with that system [9]. For example, NIST's Guide for Conducting Risk Assessments states that individual organizations are responsible for determining the level of risk that is considered acceptable [28].

Sensitivity Analysis Results

To determine what the key drivers are in the process's ability to identify and mitigate risk, manipulation of process factors through experiments was conducted. Primary factors considered were Worker Proficiency, AAR Threshold, POA&M Threshold, AAR Difficulty, and POA&M Threshold each at multiple levels. Replications were run for each of these factors at their various levels (15,000 in total), and results were compared to the baseline. While the final Total Risk Identified and Mitigated are the primary outputs of interest because they portray the amount of total system risk identified and mitigated, five other result metrics are monitored to gain a better understanding of what drives process success in terms of risk identification and mitigation. These

additional outputs of interest are Initial Risk Identified (before any rework as occurred), Final Risk Identified (after all rework has been completed), Initial Risk Mitigated, Final Risk Mitigated, and Total Number of Reworks.

Worker Proficiency

The first experiment modified Worker Proficiency to three different levels of proficiency. The baseline process conducts a random selection of any of the five proficiency levels. This random draw assumes a uniform distribution, meaning there is an equal probability (20%) of selection among the 5 possible levels. Next, it was assumed only level 2 and higher could be selected. To do this, the distribution was altered such that the probability of selecting a level 1 was 0% and the probability of selecting a level 2 was doubled (40%), while keeping the probability of selecting levels 3 and 4 the same (20%). Altering the distribution in this way simulates training level 1 employees to become level 2 proficient. In like manner, the second experiment assumed neither levels 1 or 2 could be selected. This dropped the probability of selecting a level 2 to 0% and increased the probability of selecting level 3 to 60% while the remaining levels remained 20% probable. Lastly, experiment three assumed the likely selection of a level 4 employee to be 80% with the probability of selecting level 5 remaining 20%. The results of each of these experiments were compared to the baseline to determine how increasing worker proficiency improved or worsened the seven outputs under observation.

As depicted in Figure 12, the baseline results are expressed in form of a bar graph. The black bars represent the amount of risk identified and mitigated expressed as a percentage of the total risk, while the blue bar is the sum of required reworks. Underneath this graph are the three separate experiments showing the percent improvement (green) or

degradation (red) experienced under the experimental configuration. The results indicate a fairly linear relationship between worker proficiency and 4 outputs – Initial Risk Identification, Initial Risk Mitigation, Initial Total Risk Mitigated, and Total # Reworks. It is interesting to note that competency of the workforce has little to no bearing on how much risk is identified or mitigated once the process is complete as shown in Figure 12. Increasing levels of worker competency simply improves the ability to initially find and moderate risk (depicted by green left of the dotted line), and thus reduces the amount of work it will take to produce satisfactory artifacts. This is reflected in the reduction of reworks (shown the right of the dotted line in green) as proficiency increases. Working under the assumption that the average process time for the PIT A&A is approximately 18 months, and the baseline number of reworks is about 22, each "rework" amounts to roughly 3.5 weeks of work effort. Thus, a 4% savings equates to a reduction of 3.5 weeks of work effort, a 12% savings is 9.5 weeks, and a 25 % savings equals 19.5 weeks. While the scenario of only having Advanced and Expert level employees (4-5) complete the PIT A&A is highly unlikely, for the purpose of this experiment, one can see increasing worker proficiency reduces the amount of rework required. This in turn reduces process time and improves efficiency.

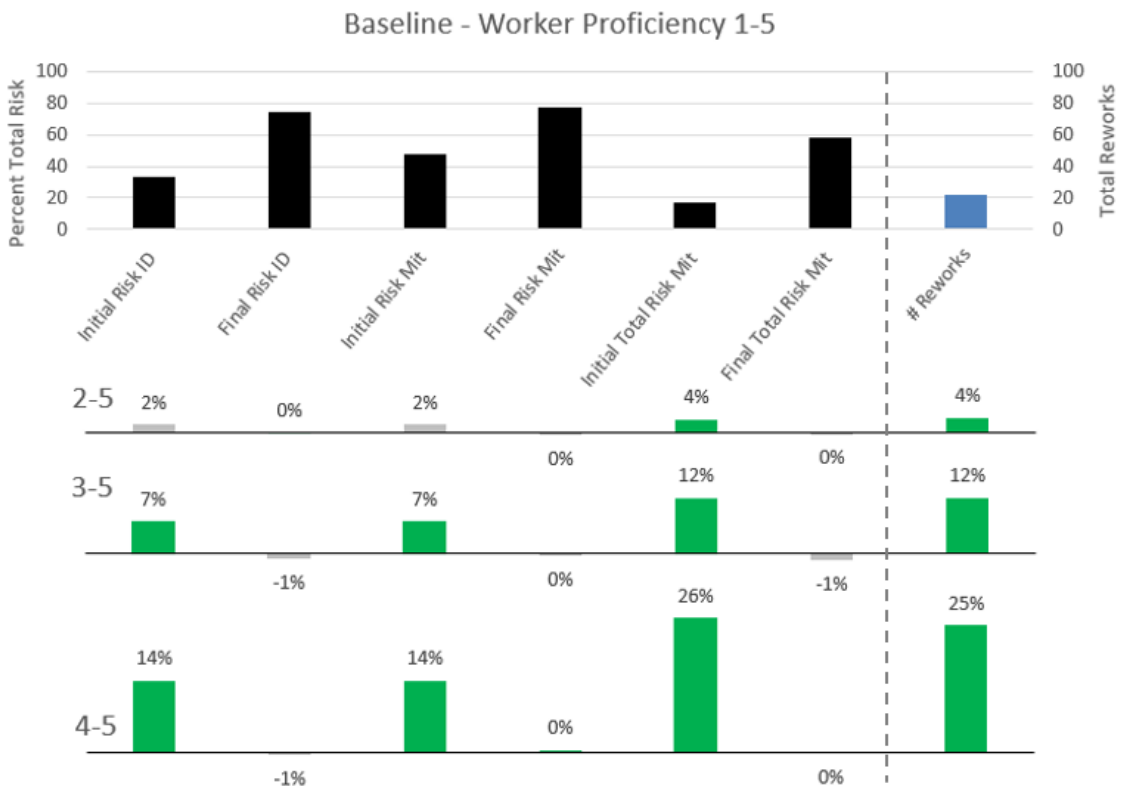


Figure 12. Sensitivity Analysis Worker Proficiency

AAR Threshold

The next experiment examined how varying the acceptance threshold of the primary risk identification artifact, i.e., the AAR, affects the observed outputs shown in Figure 13. The baseline threshold value of the AAR is 0.7 or 70%. Depicted in the black and blue bar graph, The AAR threshold was then tested in 10% increments around its baseline value. It was first reduced to 60%, then increased to 80%, and finally jumped to 90%. Compared to the baseline, green bars represent a positive response while red bars characterize negative responses.

As expected, the values below the AAR Threshold baseline have a fairly significant negative impact in the process's ability to identify risk as reflected in the Final

Risk Identification output and Total Risk Mitigated. There is a reduction of rework, however not substantial enough to counter the amount of risk left unmitigated. Conversely, as the AAR threshold increases, rework increases, where the Final Total Risk Mitigated improves just as much as Final Risk Identified. This shows a noteworthy relationship between identified and mitigated risk - the more risk the process identifies, the more risk can be mitigated. This confirms the importance of the AAR and its role in identifying risk. Because of its importance, the balance between a higher fidelity or threshold and required rework should be considered.

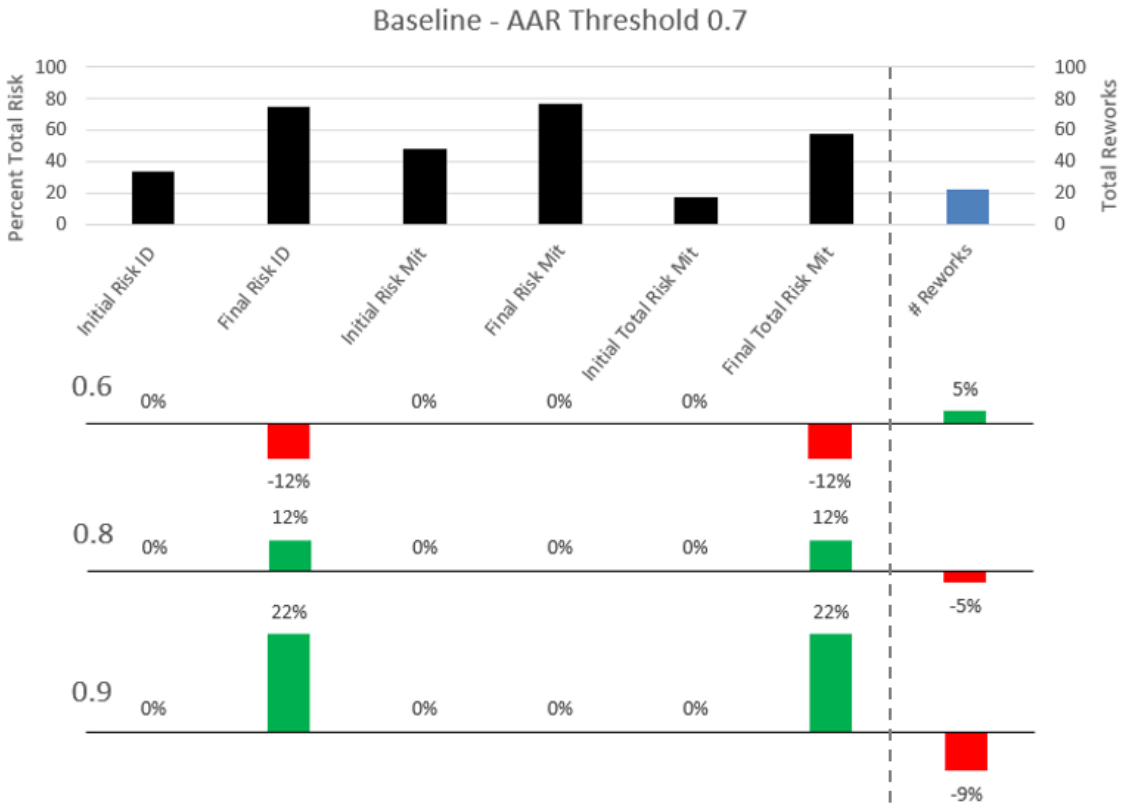


Figure 13. Sensitivity Analysis AAR Threshold

POA&M Threshold

Next, the primary artifact in risk mitigation was considered. The POA&M accounts for approximately 25% of risk mitigation as reflected in its artifact weight. Its baseline threshold value is 70% where this experiment incrementally modified POA&M threshold by 10% increments to understand impacts on the primary outputs observed. Relations compared to baseline results were analyzed in terms of percent increases (green) and decreases (red).

Similar to AAR Threshold, Figure 14 illustrates how Total Risk Mitigated decreases as POA&M Threshold is reduced. In addition, the Total Number of Reworks declines in a similar manner with threshold reduction. The drop in Total Risk Mitigated as POA&M Threshold decreases compared to drop in Total Risk Mitigated as AAR Threshold decreases is less significant. However, the reduction in POA&M Threshold results in a greater improvement with respect to amount of required rework. This indicates there is more to be gained in terms of rework reduction without compromising risk mitigation by lowering the POA&M Threshold as opposed to the AAR Threshold.

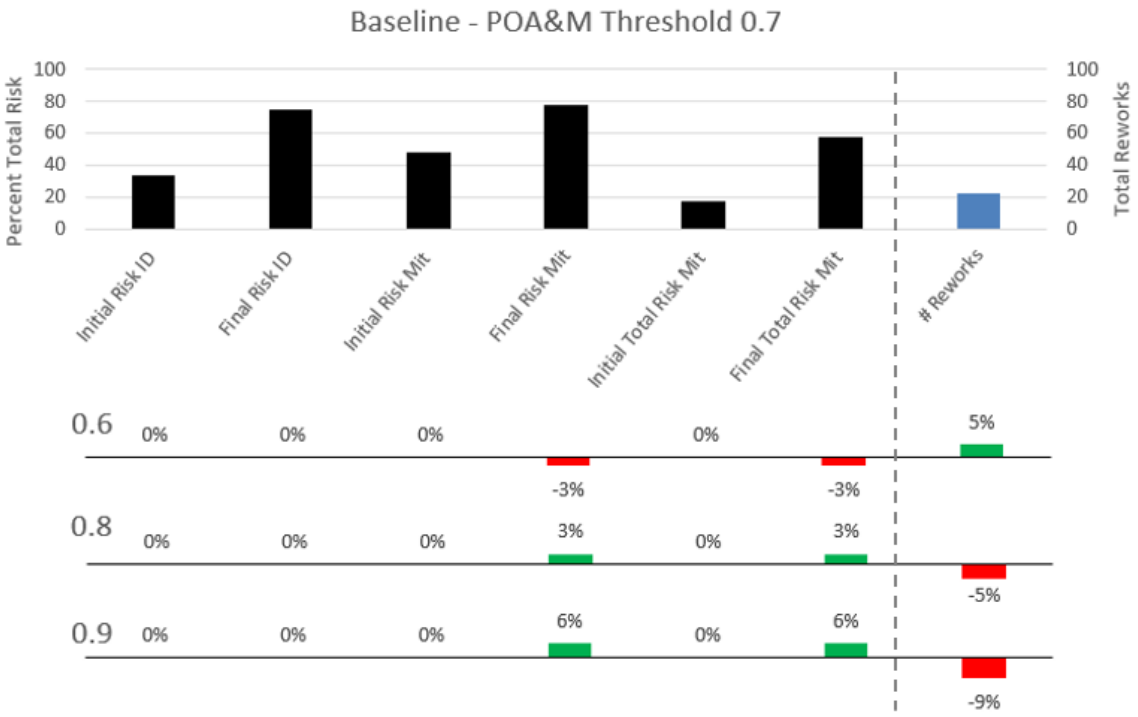


Figure 14. Sensitivity Analysis POA&M Threshold

AAR Difficulty

The next factor measured was AAR Difficulty. There is a considerable amount of rigor involved in completing the AAR. For this reason, the baseline difficulty value was appointed to be 0.45. Within this experiment, the baseline difficulty was decreased to 0.4, to 0.35 and finally to 0.3. The percent changes in each of the observed outputs was measured and compared to baseline results. Green bars characterize positive responses while red bars signify a negative response.

As shown in Figure 15, there is a substantial improvement in the Initial Risk Identified and Initial Total Risk Mitigated as AAR Difficulty is reduced as illustrated with the green bars increasing in size; however, this does not improve the end state of risk mitigation. The gray bar indicates the percent change was minor, but there is a slight

decrease in Total Risk Mitigated. There is a minor improvement (shown with the blue outlined green) in the reduction of required rework; however, the improvement comes at the expense of risk mitigation. This indicates the burden of rework is only slightly diminished by lowering AAR Difficulty. This relief of rework comes with the potential cost of reducing the amount of Total Risk Mitigated. The balance between these two outcomes should be considered when looking to simplify AAR population.

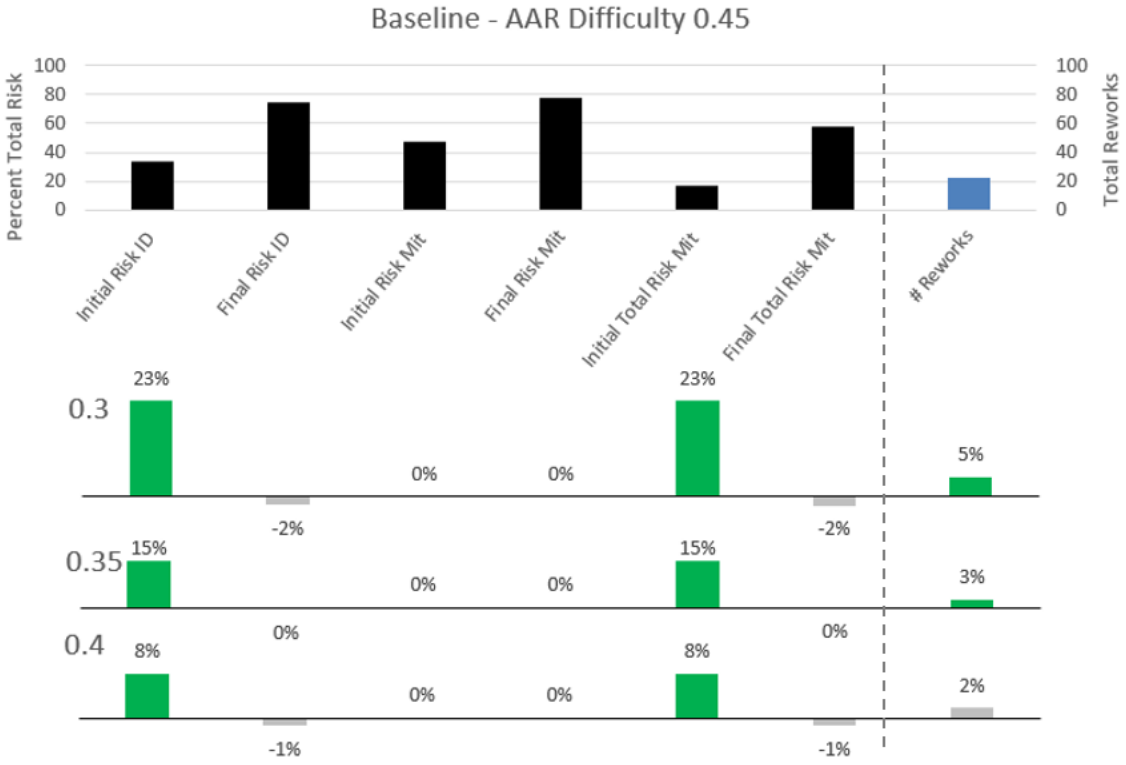


Figure 15. Sensitivity Analysis AAR Difficulty

POA&M Difficulty

This experiment looked at POA&M Difficulty. While this document is challenging to populate, the technical proficiency in its compilation is less than that of the AAR. For these reasons, the baseline difficulty value was set at 0.35. Similar to the AAR Difficulty experiment, baseline difficulty was modified by increments of 0.05. Although

it might be impractical to reduce the difficulty of such a significant artifact to such a low level, the purpose of this experiment is to provide an understanding of how much POA&M Difficulty impacts effectiveness and efficiency. This can be seen in the percent changes of each of the observed outputs in comparison to baseline results.

As expected, decreasing Artifact Difficulty reduces additional rework as shown in Figure 16. In addition, a higher difficulty lessens the ability to initially mitigate risk, but ultimately has no bearing on the Total Risk Mitigated. While these are interesting observations and further prove difficulty does not impact risk mitigation, the focus of this experiment is examining the magnitude of the changes specifically in the amount of required rework. Across the experiment, the percent changes (2%, 3%, and 4%) are basically negligible. Thus, there is very little if any benefit to changing POA&M Difficulty.

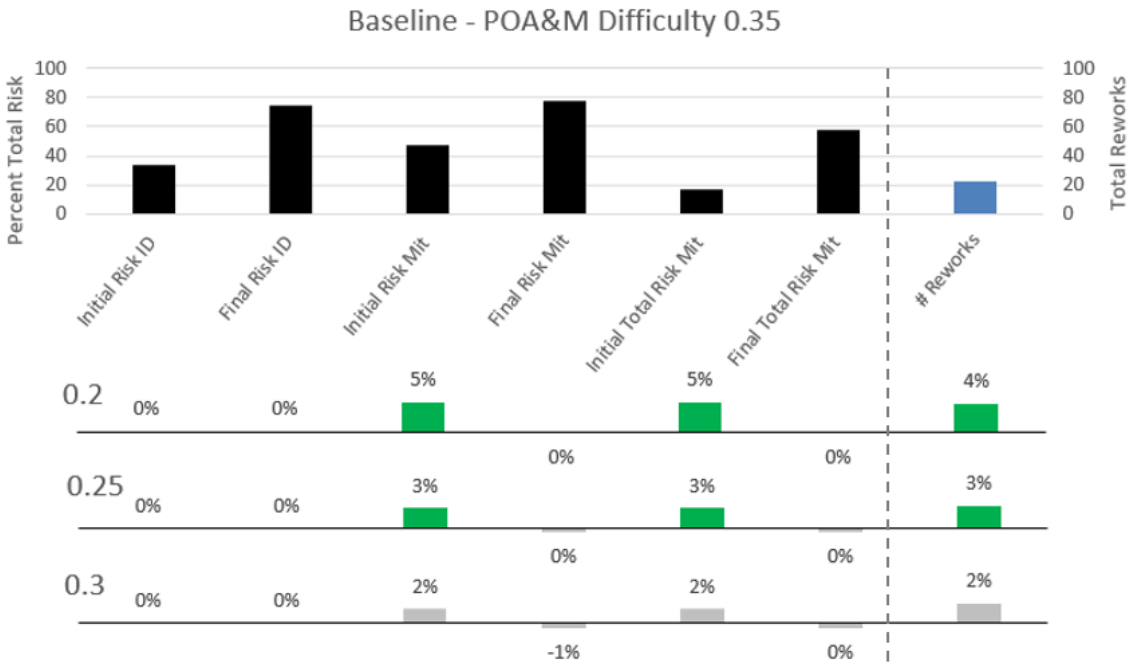


Figure 16. Sensitivity Analysis POA&M Difficulty

Design of Experiments

The objective within this designed experiment is to understand which set of factors in the PIT A&A process affects the performance the most and then determine the best levels for these factors to obtain better outcomes given realistic constraints [32]. With 5 primary factors 4 levels per factor, and an interest in 3 outcomes, a full factorial DOE would require and the calculation review of 3072 results. From the above experiments, there is a greater understanding of how individual factors impact the baseline results. Key takeaways include the fact that Worker Proficiency does not increase the amount of risk identified or mitigated. It only determines how quickly designated thresholds are met. Likewise, difficulty has little effect on risk mitigation, but can improve process time. Thresholds are primary drivers in the amount of risk the process is able to identify and mitigate.

With this understanding, a number of two-level factorial experiments are conducted to more closely study interactions between primary factors. Multiple two level factorial DOE permits more purposeful selection of interactions without the computational challenge. The selection of three designs is presented.

DOE 1

Knowing that Worker Proficiency drives rework reduction and Thresholds determine how much risk can be identified, the first design studied the interactions between Worker Proficiency, AAR Threshold, and POA&M Threshold. These interactions should start to reveal the balance between effectiveness and efficiency. As shown in Table 3, high and low levels are identified for each for each factor. It was assumed Worker Proficiency could not be improved beyond an Intermediate level employee as it is not feasible to immediately train the workforce to an Advanced level in completing the PIT A&A Process. Threshold values for the AAR and POA&M are explored at 70% as the low value and 90% as the high.

	DOE 1	***Low (-)***	***High (+)***
A	**Worker Proficiency**	1-5	3-5
B	**AAR Threshold**	0.7	0.9
C	**POA&M Threshold**	0.7	0.9

Table 3. Design of Experiments 1 Configuration

This DOE requires eight experimental configurations to achieve all possible interactions between the three factors as recorded in Table 4. Each configuration was considered as potential for possible process improvement. Because the objective is to

maximize risk identification and mitigation while minimizing rework, Final Risk Identification, Final Risk Mitigation, Final Total Risk Mitigation, and Total Number of Reworks are selected as the responses of interest. The results shown in Table 4 represent the average of the respective metric over 1000 replications. Comparing these responses to the baseline values, positive response values were colored green and negative colored red. Configurations with significant improvements in multiple primary outputs was considered for further investigation.

Working through each configuration at lower proficiency levels, configurations 3 and 4 provide a significant improvement in the Total Risk Mitigated. This increase in percent risk mitigated occurs when the AAR Threshold is high. Configuration 3 raises AAR Threshold while keeping the POA&M Threshold low, which results in an 18% increase in risk mitigation. Configuration 4 raises POA&M Threshold along with the AAR, which improves the percent increase of risk mitigated to 26%. This further demonstrates how Artifact Thresholds drive effectiveness in terms of risk mitigation. Despite the increases in Total Risk Mitigated, options 3 and 4 reduce efficiency with an increase in required rework. Knowing Worker Proficiency can improve efficiency, Configurations 7 and 8 (where Thresholds are high) should reveal a balance of effectiveness and efficiency. Although option 8 necessitates added rework, this is more balanced than option 4. With a 19% increase in risk mitigated and a 7% improvement in Total Reworks, Configuration 7 (with high proficiency, high AAR Threshold, but a low POA&M Threshold) proves to be the ideal option from this design. This indicates it may not be advantageous to max all artifact thresholds. In order to keep process time low but

effectiveness high, thresholds should only be high for those artifacts that are key in risk mitigation.

DOE 1	*A*	*B*	*C*	***Final Risk ID***	***Final Risk Mit***	***Total Risk Mit***	***Total # Reworks***
1	-1	-1	-1	75.09	**74.73**	**56.12**	**21.53**
2	-1	-1	1	75.08	**79.61**	**59.78**	**23.47**
3	-1	1	-1	**91.88**	**74.65**	**68.58**	**23.54**
4	-1	1	1	**91.96**	**79.47**	**73.09**	**25.37**
5	1	-1	-1	74.05	**74.43**	**55.12**	**18.68**
6	1	-1	1	74.10	**79.29**	58.75	**20.82**
7	1	1	-1	**92.35**	**74.48**	**68.79**	**20.69**
8	1	1	1	**92.21**	**79.38**	**73.20**	**23.01**
Baseline				**75.05**	**77.22**	**57.96**	**22.21**

Table 4. Design of Experiments 1 Results

DOE 2

With the understanding that Worker Proficiency and Artifact Difficulty are the forcing functions behind reduced rework, the second design considered Worker Proficiency, AAR Difficulty and POA&M Difficulty. The goal with these interactions (shown in Table 5) is to determine how much efficiency can improve in terms of rework while monitoring the effect on mitigated risk. Because the POA&M requires a considerable amount of effort, it did not seem practical to test a Difficulty value less than 0.3. Similarly, a lower bound below 0.35 for the AAR Difficulty appeared unrealistic due to the intricacy of this artifact.

	DOE 2	***Low (-)***	***High (+)***
A	**Worker Proficiency**	1-5	3-5
B	**AAR Difficulty**	0.4	0.45
C	**POA&M Difficulty**	0.3	0.35

Table 5. Design of Experiments 2 Configuration

Results from the eight configurations in DOE 2 are presented in Table 6. As predicted, there was very little movement in the amount of risk identified and mitigated, but there were reasonable improvements in the amount of rework. Although the movements in Total Risk Mitigated were slight, they were all in the wrong direction. This shows how risk mitigation is compromised with the reduction of Artifact Difficulty. Configuration 5 experiences the most significant drop in Total Reworks, which equates to a 17.5 week cut to the process time. However, this time savings comes with a reduction in mitigated risk. While none of these configurations are ideal, there is value in seeing the tradeoffs between risk mitigation and required rework as Artifact Difficulty declines. With this understanding, increasing Thresholds could potentially counter the adverse effects on risk mitigation at lower Difficulty levels while maintaining improvements in required rework.

DOE 2	*A*	*B*	*C*	*Final Risk ID*	*Final Risk Mit*	*Total Risk Mit*	*Total # Reworks*
1	-1	-1	-1	74.06	**74.28**	**55.01**	**20.56**
2	-1	-1	1	74.02	**74.67**	**55.27**	**21.21**
3	-1	1	-1	75.04	**74.24**	**55.72**	**21.41**
4	-1	1	1	75.22	**74.68**	**56.18**	**21.02**
5	1	-1	-1	73.99	**73.73**	**54.56**	**17.18**
6	1	-1	1	74.00	**74.43**	**55.08**	**18.22**
7	1	1	-1	74.17	**73.66**	**54.63**	**18.20**
8	1	1	1	74.39	**74.40**	**55.36**	**18.51**
Baseline				**75.05**	**77.22**	**57.96**	**22.21**

Table 6. Design of Experiments 2 Results

DOE 3

Building off results from DOE 2, interactions between Worker Proficiency, Threshold, and Difficulty are explored. Changing one factor at a time in the sensitivity analysis revealed the significant influence of the AAR within the PIT A&A Process and suggested further examination. This last design examined only the AAR and it is interaction with Worker Proficiency levels. This was done to look for further confirmation of the AAR's importance and to understand how proficiency levels and difficulty levels impact the rework required for this highly technical document. The High/Low levels of these factors are seen in Table 7.

	DOE 3	***Low (-)***	***High (+)***
A	**Worker Proficiency**	1-5	3-5
B	**AAR Threshold**	0.7	0.9
C	**AAR Difficulty**	0.4	0.45

Table 7. Design of Experiments 3 Configuration

Results for this design are offered below in Table 8. Configuration 4 shows the amount of risk that can be mitigated if AAR Difficulty is low and Threshold is high but Workforce Proficiency is constrained. Configuration 3 should demonstrate the improvement from Configuration 4 with respect to required rework if Difficulty is lessened. However, the improvement from 23.19 reworks to 23.15 is negligible. The same improvements can be observed from Configuration 8 to Configuration 7 where proficiency is higher. With a higher competency, the improvement in required reworks is more substantial. The significant increase in Total Risk Mitigated where Threshold is high further proves the importance of the AAR with respect to risk mitigation. This design also reveals that improvements in terms of required rework with a reduction in Artifact Difficulty are only realized if Worker Proficiency is high.

DOE 3	*A*	*B*	*C*	***Final Risk ID***	***Final Risk Mit***	***Total Risk Mit***	***Total # Reworks***
1	-1	-1	-1	74.25	**74.65**	**55.43**	21.65
2	-1	-1	1	75.24	**74.62**	**56.15**	**21.07**
3	-1	1	-1	**91.52**	**74.71**	**68.37**	**23.15**
4	-1	1	1	**92.00**	**74.69**	**68.72**	**23.19**
5	1	-1	-1	74.02	**74.37**	**55.05**	**18.70**
6	1	-1	1	74.32	**74.47**	**55.36**	**18.07**
7	1	1	-1	**91.84**	**74.49**	**68.42**	**19.43**
8	1	1	1	**92.25**	**74.44**	**68.67**	**20.74**
Baseline				**75.05**	**77.22**	**57.96**	**22.21**

Table 8. Design of Experiments 3 Results

Recommended Configuration Options

Analyzing these interactions provides a greater understanding of the tradeoffs between risk mitigation and required rework. In addition, consideration has to be given to

the extent to which Worker Proficiency can improve, Thresholds are raised, and Difficulties are lowered. In many cases, the cost of associated with these changes may be too great or unrealistic. For example, in order to increase Workforce Proficiency, 100% of employees must be trained to assess cyber risks of MWSs. While concentrated specialized training to achieve specific tasks can increase workforce confidence in their ability to conduct effective assessments, experience is difficult to accelerate. Artifact Thresholds can be increased with systematic structured quality assurance and overhead review, and Artifact Difficulty can be reduced by creating templates and providing examples. Some of these investments are more plausible than others. For these reasons, recommendations come in the form of a solution trade space rather than a single point solution. Keeping this in mind, the following options present potential candidate configurations to maximize Risk Identification and Mitigation while minimizing Rework.

	#1 Ideal	*#2 Workforce Limited*	*#3 Minimum Reworks*	*#4 Most Feasible*
Worker Proficiency	3-5	1-5	3-5	2-5
AAR Threshold	0.9	0.9	0.8	0.8
AAR Difficulty	0.4	0.4	0.4	0.4
POA&M Threshold	0.9	0.9	0.7	0.8
POA&M Difficulty	0.3	0.3	0.3	0.3

Table 9. Configuration Options

Variance between these options is demonstrated in Figures 17 and 18 where the variance of both the Total Risk Mitigated and the Total Number of Reworks for each option is compared to that of the Baseline (represented by the blue boxes). The amount of variance is characterized by the size of the colored boxes. In the comparison of two

options, if no overlap between the boxes exists, there is evidence the median values are significantly different or the centers are statistically significant [33].

In order to confirm statistical significance, a two-tailed, paired t-test was conducted between each option and the baseline. A two-tailed test is conducted when there is possibility of either a positive or negative relationship between data sets. In comparing two variables, t-tests return p-values which indicate the percent chance the results happened by accident. The larger the p-value, the less likely the two variables under observation are different. The t-test conducted assumed a significance value (alpha) of 0.05. With a null hypothesis that there is no difference between the two options, a calculated p-value of less the 0.05 would result in a rejection of the null. Thus, there is evidence of statistical significance that one option is different (better or worse) than the other.

Option #1 is the Ideal scenario – Worker Proficiency is higher, Thresholds are maxed and Difficulties are low. The Ideal Option colored in red shows about a 24% increase in Risk Mitigation with a much smaller variance. In Figure 17, there appears to be no overlap between Option 1 and the Baseline indicating there is statistical significance that Option 1 is better in terms of risk mitigation. The p-value confirms this evidence since it is far less than the significance value of 0.05. Comparing the average number of reworks required, the Ideal option indicate a reduction in process time by approximately 3 weeks. This difference is not statistically significant which is reflected in a larger p-value as seen in Figure 18. With this p-value, the null hypothesis is assumed to be true - there is no difference in required rework. Weighing the cost to implement changes to achieve these levels may be too great for the reward. Training all level 1 and 2

employees to be have an Intermediate proficiency level in and of itself may be challenging, but not unreasonable. Accomplishing this along with reducing Artifact Difficulty and raising Thresholds to levels in this configuration would be ideal, but most likely infeasible.

Option #2 is Workforce Limited and it recognizes the challenge in training the workforce enough to achieve higher proficiency. Still, it assumes Thresholds can be maxed and Difficulty can be decreased. Because of the reduced proficiency there is larger variability in the Option #2 as seen in the purple box within Figure 17. The variability is not sizable enough to cause overlap with the Baseline, which leads to the same conclusions as the previous option – there is statistical significance that Option #2 is superior to the Baseline in terms of risk mitigation. Again, the lower p-value confirms this significant increase in risk mitigation. The p-value for required rework (shown in Figure 18) also rejects the null hypothesis indicating a statistically significant difference. However, this difference is in the wrong direction with an increase in rework. Although this option achieves a much higher percentage of Total Risk Mitigated, this configuration only solves half the problem. The evidence of a statistically significant increase in Total Reworks (equating to an average of 12 weeks), indicates this option does not provide the appropriate balance between effectiveness and efficiency.

Option #3 or Minimum Reworks attempts to reduce the required rework while maintaining a substantial amount of risk mitigated. This is done by keeping the Thresholds high (but not at max value), reducing Difficulty, and improving Worker Proficiency. While this option (represented with the green boxes in Figures 17 and 18) does not achieve the same amount of Risk Mitigated, there is still an improvement from

the baseline with respect to both risk mitigation and required reworks. P-values for both Risk Mitigated and Total Reworks are less than 0.05, proving statistical significance of these improvements. Because Threshold values are not as high as Option #1, this is a more viable configuration. However, there is still the challenge of training the workforce to meet the competency of a Level 3 employee.

Lastly, Option #4 (Most Feasible) attempts to find the middle ground in terms of Worker Proficiency. Where achieving the competency of a level 3 employee would require strict training requirements, educating the workforce enough to be level 2 proficient would be more reasonable. Thus, Worker Proficiency under this configuration assumes levels 2-5 while Thresholds are slightly increased, and Difficulty is decreased. Within the comparison of this option (shown in orange) and the Baseline, the null hypothesis is rejected with respect risk mitigation due to its very low p-value indicating there is a significant difference between the two. This verifies the statistical significance of the improvement in risk mitigation under Option #4. On average, this enhancement consists of a 13% increase in the amount of risk mitigated. The null hypothesis in comparing required reworks for this option, however, is accepted with a p-value of 0.743. There is no statistical evidence of a significant difference between the Baseline and Option #4. In fact, the average number of reworks is approximately the same. Weighing the investments needed to achieve levels within this configuration with the performance improvements, this option was considered to be the Most Feasible.

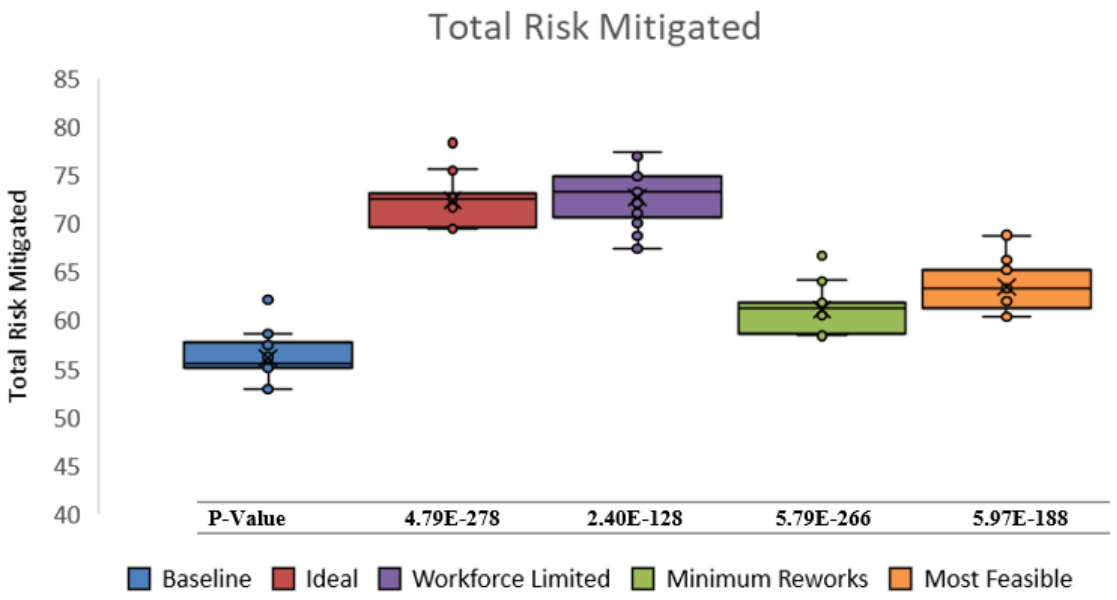


Figure 17. Risk Mitigation Statistical Significance Between Options

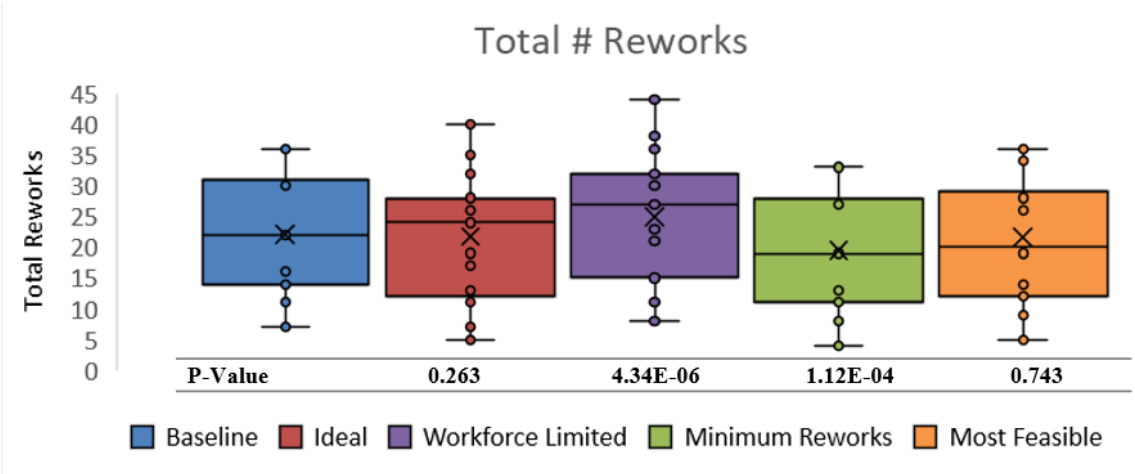


Figure 18. Rework Statistical Significance Between Options

Summary

After reviewing the tradeoffs between these four recommended options and comparing results to the baseline, Option #4 proved to be most feasible and provided statistically significant improvements to risk mitigation without adding any required rework. The four options presented were configured based of critical understanding gained through a series of experiments presented in this chapter. First, the chapter established baseline results of model outputs under a configuration representative of the actual PIT A&A Process. Next, sensitivity analysis changed one factor at a time and observed changes in from the baseline to determine critical drivers in process effectiveness and efficiency. Interactions between those critical factors were further examined in three different experimental designs. From these experimental configurations, the four recommended options were designed, simulated and analyzed.

V. Conclusions and Recommendations

Chapter Overview

Within this final chapter, conclusions and recommendations are presented by reflecting on the problem space and describing how this research can improve the state of cybersecurity within MWSs by understanding and focusing on key drivers in risk identification and mitigation. This is done by working through each investigative question that guided the development and testing of the PIT A&A Process model. Simulation results and recommendations for process improvement are summarized. Research significance is explained and potential future research efforts are suggested.

Conclusions of Research

With the rising concern that military systems are vulnerable to exploitation and offensive attack through cyberspace, it is crucial that DoD organizations effectively identify and mitigate risks. Despite security requirements levied upon system owners and developers, compliance has not resulted in more secure and defensible weapon systems. This research was conducted with the intent to help shift DoD emphasis from compliance to mission assurance by understanding and focusing the PIT A&A Process on activities which enhance the identification and mitigation of cyber risks. It should be noted that results and conclusions drawn from model performance are influenced by assumptions outlined in Chapter III.

- **Research Question #1**

The first question framed the research effort around the effectiveness of the PIT A&A Process. “How can the PIT A&A Process be studied using modeling and

simulation?" In order to answer this question, the PIT A&A Process must first be well understood. Chapter II offers history behind its development as well as detailed instruction on how each step should be carried out. This familiarity with the process establishes the foundation upon which all modeling assumptions are built. With this understanding the next step in answering this question can be taken – developing a simulation model. Presented in Chapter III, the development of a stochastic model using 12 Steps is detailed. The in-depth study of the PIT A&A Process led to greater awareness of factor criticality and how each factor affects subsequent steps within the process. Based on this understanding, assigned values associated with these factors are assumed. Next, model conceptualization, development, verification and validation ensured not only proper function of the model, but also proper representation of the PIT A&A Process. Once model development is complete, controllable experimental factors are manipulated to better understand process dependencies, effectiveness, and efficiency. In short, the PIT A&A Process can be studied using modeling and simulation by first understanding the process and its critical factors, and then following a methodical approach to develop a model representative of the process.

- **Research Question #2**

With this model in place, the second research question can be addressed: "How can a baseline for measuring the effectiveness of the PIT A&A Process be established?" Based on the model, a baseline configuration of controllable factors was selected to represent the PIT A&A Process. Values were chosen for this configuration based on interpretation of existing process documentation as shown in Table 2 within Chapter III.

Model results provide a baseline (or a reasonable estimation) of process effectiveness in terms of how much risk the PIT A&A Process identifies and mitigates through multiple reworks. These baseline results are presented in Chapter IV and are illustrated in Figure 11. Based on my assumptions, these results indicate that the PIT A&A Process identifies 75% and mitigates 57% of the Total System Risk. However, this measurement of effectiveness is subjective. What amount of identified/mitigated risk is effective or sufficient? This is dependent upon mission assurance requirements associated with the system under assessment. For example, mitigation of 57% may be effective for an AIM-9 Missile, but not for an F22. It is ultimately the individual organization's responsibility to determine what is satisfactory. Although simulation results provide a quantitative measure of process effectiveness, they still are subjective.

- **Research Question #3**

The next research question is: "How does workforce proficiency, accuracy thresholds of security artifacts, and difficulty of artifact completion impact the PIT A&A Process's ability to identify and mitigate cybersecurity risks?" This question prompted investigation of the model to better understand what factors influence success in terms of risk identification and mitigation detailed in Chapter IV of this work. Several factors were explored to understand impacts on the process outputs and then key interactions were more thoroughly studied in three experiments. From these experiments, important relationships associated with these factors were discovered. First, Worker Proficiency impacts how rapidly thresholds are achieved, but does not increase the amount of risk identified or mitigated. This means training personnel to become more proficient in

conducting cybersecurity assessments of MWSs will not improve the amount of risk identified or mitigated; however, it will increase the accuracy of key artifacts so less rework is required. Educating employees only accelerates the achievement of acceptable Artifact Accuracy, thus improving efficiency.

Similarly, Artifact Difficulty can improve process time, but has little effect on risk mitigated. This indicates efforts to reduce Artifact Difficulty (e.g., providing templates, examples of accurate artifacts, and training) does not increase effective risk mitigation but will decrease the time and effort required to produce a satisfactory authorization package. Lastly, Artifact Thresholds are principal drivers in the amount of risk the PIT A&A Process is able to identify and mitigate. Thresholds are managed by how well authorization authorities and highly trained personnel evaluate Artifact Accuracy. Because authorization authorities control Artifact Thresholds, the amount of risk identified and mitigated through the PIT A&A Process lies in their hands. However, measures can be taken to improve this evaluation process such as implementing structured quality assurance or overhead review.

- **Research Question #4**

With an understanding of controllable factors and process outputs, the final research question can be addressed: "How can the PIT A&A Process maximize risk mitigation while minimizing artifact rework?" With insight gained from designed experiments, Chapter IV presents four potential model configurations. In terms of Total Risk Mitigated, all options demonstrated statistically significant improvements over the

baseline as shown in Figure 17 with option #1 @ ~28%, #2 @ ~29%, #3 @~9%, and #4 @~13%.

Regarding the amount of rework required, only two options (#2 and #3) showed statistical significance – while Option #3 notably reduced rework compared to baseline results, Option #2 displayed significant amounts of rework added. These options are not necessarily point solutions, but rather illustrations of the tradeoffs associated with possible changes to the PIT A&A Process. Although it may not be unreasonable to consider increasing workforce proficiency to level 3 or higher, there would be serious challenges in implementing this requirement. Educating an employee with a level 1 proficiency to be as competent as a level 3 would require a significant amount of detailed training specific to respective systems as well as effective risk management. Such a curriculum has yet to be developed. For this reason, Options #1 and #3 did not appear to be reasonable achievements. Option #2 does not require any changes to Worker Proficiency and still achieves a much higher percentage of Total Risk Mitigated. However, the amount of required rework increases along with risk mitigated. In fact, there is a statistically significant increase in Total Reworks (shown in Figure 18), which does not offer the appropriate balance between effectiveness and efficiency. The best solution providing the ideal balance between maximizing risk mitigation while minimizing reworks is Option #3. However, due to limitations in developing high levels of proficiency, this option was deemed impractical.

Building from the configuration of Option #3, Option #4 assumes level 2 proficiency levels can be attained through basic training requirements. This option achieves statistically significant improvements in risk mitigation without any added

rework compared to baseline results. This option is considered most feasible and still provides a balance between maximizing risk mitigation while minimizing rework.

For these reasons, it is recommended the PIT A&A Process implement changes to accomplish levels of Worker Proficiency, Artifact Difficulties, and Artifact Thresholds that closely reflect those of Option #4. Again, this configuration entails Worker Proficiency levels 2-5, an AAR Difficulty of 0.4 with a POA&M Difficulty of 0.3, and Threshold values for both AAR and POA&M set at 0.8. To do this, training requirements need to be established mandating employees achieve increased competency before engaging in risk assessment activities. This training will increase Worker Proficiency and assist in reducing Artifact Difficulty.

Although requirements have not been formally established, training courses have been developed to help inform personnel on about the RMF and PIT A&A Process. No formal training has been created to guide employees through successful compilation of required artifacts. Additionally, templates for all critical artifacts should be provided along with detailed examples to produce the highest accuracy possible. AFLCMC has developed draft templates for a number of artifacts. Finalizing templates for all artifacts along with detailed examples would simplify successful artifact completion and reduce required rework. Lastly, authorization authorities should implement systematic structured quality assurance to ensure Artifact Accuracy levels reach higher Thresholds.

Significance of Research

The significance of this research lies not in the model's ability to specifically quantify effectiveness of the PIT A&A Process, but to identify key drivers in the risk

identification and risk mitigation process. Awareness of the key drivers in process effectiveness and efficiency provides a roadmap for organizations to improve risk assessments in order to develop, test, and assure more cyber-secure systems. Knowing how effective this process can be at mitigating cyber risks should promote a better understanding of the process with the goal of making MWSs more secure.

Successfully showing the interchanges between process effectiveness and efficiency by manipulating controllable factors demonstrates how effective modeling and simulation can be in studying processes. This simulation not only facilitates better understanding how to more effectively accomplish the PIT A&A Process from a risk identification and mitigation perspective, but also where process efficiencies can be made in terms of reducing required rework.

Recommendations for Future Research

No simulation model is a completely accurate representation of the system or process. Still, the model presented in this work is assumed to be sufficient to answer the questions guiding this research. However, a number of actions can be taken to refine the model and thus increase confidence in simulation results.

First, it is recommended data be recorded to more accurately monitor process performance. The amount of time a particular artifact is in development or under revision should be tracked and compared to establish timelines. Moreover, these timelines should be tied to various classes of systems undergoing assessment (e.g., complex systems, simple systems, etc.). The number of iterations or reworks required before acceptable artifact accuracy is achieved should also be monitored for each artifact. This would

establish a factual understanding of where the workforce is lacking proficiency and what training is needed to improve the process.

Next, the assumptions made within this work may be an oversimplified means to measure risk; however, the ability to quantify and measure risk would prove to be very valuable. It is recommended a more precise way to define risk be established to determine quantified risk scoring. This would provide clarity to a very subjective space which would promote better understanding and thus improve critical decision making. Because of the increasing concern of cyber attack and vulnerability exploitation along with the rising importance of fielding resilient MWSs, it is recommended a team of independent process engineers further study the effectiveness and efficiency of the PIT A&A Process.

Bibliography

[1] B. Brenner, "WannaCry: The Ransomware Worm That Didn't Arrive on a Phishing Hook," *naked security*, 2017. [Online]. Available: https://nakedsecurity.sophos.com/2017/05/17/wannacry-the-ransomware-worm-that-didnt-arrive-on-a-phishing-hook/. [Accessed: 19-Jul-2017].

[2] Executive Office of the President (DHS), "Homeland security presidential directive 7: Critical infrastructure identification, prioritization, and protection," *Washington, DC White House*, 2003.

[3] S. Laswon, "Just How Big is the Cyber Threat to the Department of Defense?," *Forbes.com*, 2010. [Online]. Available: https://www.forbes.com/sites/firewall/2010/06/04/just-how-big-is-the-cyber-threat-to-dod/#1b71bbd746b3. [Accessed: 19-Jul-2017].

[4] K. Baldwin, P. R. Popick, J. F. Miller, and J. Goodnight, "The United States Department of Defense Revitalization of System Security Engineering Through Program Protection," in *Systems Conference (SysCon), 2012 International*, 2012.

[5] J. Gould, "To Find Cyber Flaws in Weapon Systems, DoD Will Move Millions," *Defense News*, 01-Sep-2016. [Online]. Available: https://www.defensenews.com/2016/09/01/to-find-cyber-flaws-in-weapon-systems-dod-will-move-millions/.

[6] P. Welsh, "Air Force looks to ensure cyber resiliency in weapons systems through new office," 2017. [Online]. Available: http://www.wpafb.af.mil/News/Article-Display/Article/1040728/air-force-looks-to-ensure-cyber-resiliency-in-weapons-systems-through-new-office/.

[7] D. Commons, “DITSCAP - DoD’s Answer to Secure Systems,” *SANS Inst.*, no. GSEC Practical Assignment Version 1.2e, 2002.

[8] Nist, “Guide for Applying the Risk Management Framework to Federal Information Systems,” *NIST Spec. Publ. 800-37*, vol. Rev 1, no. February, p. 93, 2010.

[9] AFLCMC/EZA/EZB/EZC, “Air Force Life Cycle Management Center Standard Process For Cybersecurity Assessment and Authorization.” 2017.

[10] D. Snyder, J. D. Powers, E. Bodine-baron, B. Fox, L. Kendrick, and M. H. Powell, *Improving the Cybersecurity of U . S . Air Force Military Systems Throughout Their Life Cycles*. 2015.

[11] J. April, F. Glover, J. P. Kelly, and M. Laguna, “Practical Introduction to Simulation Optimization,” *Proc. Winter Simul. Conf.*, pp. 71–78, 2003.

[12] A. Singer, P.w, Friedman, “Cybersecurity and Cyberwar,” *Igarss 2014*, no. 1, pp. 1–5, 2014.

[13] A. F. P. I. W. Group, *IA PIT Guidebook*. .

[14] DODI, “Cybersecurity,” vol. 8500.01, no. 8500, pp. 1–59, 2014.

[15] Department of Defense Chief Information, “DoD Strategy for Defending Networks, Systems, and Data,” 2013.

[16] United States Air Force, “Cyber Vision 2025.” 2012.

[17] Office of Assistant Secretary of Defense Memorandum, “The Defense Information Systems Security Program (DISSP).”

[18] *Title 44 U.S. Code Chapter 35, Subchapter III - Information Security*. 2014.

[19] DODI, “DoD Information Technology Security Certification and Accreditation Process (DITSCAP),” vol. 5200.40, no. 5200, pp. 1–68, 1997.

[20] “Transitioning from DIACAP to RMF,” *PhoenixTS,* 2013. [Online]. Available: https://phoenixts.com/blog/diacap-vs-rmf/. [Accessed: 19-Nov-2017].

[21] D. of Defense, “Risk Management Framework for DoD IT,” no. 8510, pp. 1–47, 2014.

[22] DODI, “Information Assurance (IA) Implementation,” vol. 8500.2, no. 8500, pp. 1–102, 2003.

[23] G. Fleener, M. Mayor, and C. Zou, “Risk Management Framework (RMF) Transition Impacts in Training Simulation Systems,” no. 15009, pp. 1–10, 2015.

[24] Air Force Platform IT Working Group, “IA PIT Guidebook,” 2013.

[25] H. J. Van Norman, “Cybersecurity and Risk Management Framework Agile ATO Assessment & Authorization,” in *Cybersecurity and Risk Management Framework - EZA 240,* 2016.

[26] National Institute of Standards and Technology (NIST), “SP 800-53 Rev.4 - Security and Privacy Controls for Federal Information Systems and Organizations,” *Natl. Inst. Stand. Technol. - Spec. Publ.,* vol. 800–53, pp. 1–460, 2014.

[27] J. Banks, B. L. Nelson, J. S. Carson, and D. M. Nicol, “Discrete-Event System Simulation,” *PrenticeHall Int. Ser. Ind. Syst. Eng.*, p. 640, 2010.

[28] NIST, “Guide for conducting risk assessments,” *NIST Spec. Publ.,* no. September, p. 95, 2012.

[29] U.S. Office of Personnel Management, “Proficiency Levels for Leadership Competencies.” [Online]. Available: https://www.opm.gov/policy-data-oversight/assessment-and-selection/competencies/proficiency-levels-for-leadership-competencies.pdf. [Accessed: 20-Nov-2017].

[30] R. G. Sargent, “Verification And Validation of Simulation Models,” in *Simulation*, 2010, no. 2001, pp. 135–150.

[31] G. Hurlburt, “‘Good Enough’ Security: The Best We’ll Ever Have,” *Computer (Long. Beach. Calif).*, vol. 49, no. 7, pp. 98–101, 2016.

[32] A. Jiju, “Design of experiments for engineers and scientists,” *With Appl. To Eng. Sci.*, vol. 37, p. 728, 2003.

[33] “Box Plots,” *NCSS Statistical Software*, 2010. [Online]. Available: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Box_Plots.pdf. [Accessed: 02-Feb-2018].

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
22-02-2018	Master's Thesis	August 2016 – March 2018

TITLE AND SUBTITLE

Cybersecurity Assessment and Mitigation Stochastic Model

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Davis, Matthew W., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENV)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENV-MS-18-M-194

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AGENCY Cyber Security Engineering Group
ADDRESS 2145 Monahan Way, WPAFB, OH 45433-7017
PHONE and EMAIL (937) 986-6715
harrell.van_norman@us.af.mil
ATTN: Harrell Van Norman

10. SPONSOR/MONITOR'S ACRONYM(S)
AFLCMC/EZAS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

With numerous cybersecurity incidents and vulnerability concerns in an increasingly contested cyber warfighting environment, the Department of Defense (DoD) has mandated cybersecurity assessment and authorization of all major weapon systems (MWS) before their use. In response to this direction, the Air Force Life Cycle Management Center (AFLCMC) created the Platform Information Technology Assessment and Authorization (PIT A&A) Process. Modeled after the NIST Risk Management Framework (RMF), the process applies a risk-based approach to cybersecurity with the goal of identifying risks and mitigating vulnerabilities in MWS. Within this work, a stochastic model of the PIT A&A Process is presented with an emphasis on understanding how the complexity of systems, accuracy of security artifacts, and workforce proficiency impacts the ability to effectively mitigate cybersecurity risks.

15. SUBJECT TERMS
Cybersecurity; Platform Information Technology; Assessment and Authorization; Risk Management Framework; Modeling and Simulation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Lt Col Logan O. Mailloux, AFIT/ENV
U	**U**	**U**	**UU**	85	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 3348 (logan.mailloux@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18