Application of Capability-Based Cyber Risk Assessment Methodology to a Space System

Martha McNeil; Thomas Llansó; Dallas Pearson
The Johns Hopkins University Applied Physics Laboratory
Laurel, Maryland, USA
martha.mcneil@jhuapl.edu; thomas.llanso@jhuapl.edu; dallas.pearson@jhuapl.edu

ABSTRACT

Despite more than a decade of heightened focus on cybersecurity, cyber threats remain an ongoing and growing concern [1]–[3]. Stakeholders often perform cyber risk assessments in order to understand potential mission impacts due to cyber threats. One common approach to cyber risk assessment is event-based analysis which usually considers adverse events, effects, and paths through a system, then estimates the effort/likelihood and mission impact of such attacks. When conducted manually, this type of approach is labor-intensive, subjective, and does not scale well to complex systems. As an alternative, we present an automated capability-based risk assessment approach, compare it to manual event-based analysis approaches, describe its application to a notional space system ground segment, and discuss the results.

CCS CONCEPTS

• Security and privacy~Systems security

KEYWORDS

Cyber; security; risk assessment

ACM Reference format:

M. McNeil, T. Llanso, D. Pearson. 2018. Application of Capability-Based Cyber Risk Assessment Methodology to a Space System. In *Proceedings of Hot Topics in the Science of Security, Raleigh, NC, USA , April 2018 (HoTSoS '18), 10 pages.* DOI: 10.1145/3190619.3190644

1 INTRODUCTION

Despite more than a decade of heightened focus on cybersecurity, cyber threats remain an ongoing and growing concern [1]–[3]. Both the rate of cyberattacks against information systems and the sophistication of attackers continue to grow [4]. In 2014, the Center for Strategic and International Studies identified cybercrime as a "growth industry" costing the global economy up to $575 billion annually [5]. More recently, Symantec reported that "Cyber attackers revealed new levels of ambition in 2016, a year marked by extraordinary attacks, including multi-million-dollar virtual bank heists, overt attempts to disrupt the US electoral process by state-sponsored groups, and some of the biggest distributed denial of service (DDoS) attacks on record powered by a botnet of Internet of Things (IoT) devices." [6]

Cyber systems are ubiquitous in all aspects of society, including traditional information systems (IS), critical infrastructure, medical, military, communications, and home and personal systems. Breaches to cyber systems continue to be front-page news [7] and the threat continues to evolve. The cyber technology and knowledge necessary to solve this problem are also continually evolving and expanding. A report by Cybersecurity Ventures predicts that "global spending on cybersecurity products and services will exceed $1 trillion cumulatively over the next five years, from 2017 to 2021" a 35-fold increase over 5 years [8].

To understand possible mission impacts due to cyber threats, stakeholders, including mission owners, must first assess their reliance on cyber-enabled systems, particularly the risks faced due to potential cyberattacks and other failures in these systems, including design flaws and operator errors. One approach to cyber risk assessment is attack- or event-based analysis that involves attempting to enumerate vulnerabilities of and attack paths through a system, followed by expert-based scoring to estimate the event likelihood or level of adversary effort and mission impact of attacks against the system. When performed with limited automation, this analysis can be subjective and may not scale to complex systems.

As an alternative, we developed an automated, capability-based risk assessment approach and tool called BluGen [9] that combines system and mission knowledge from system stakeholders with a database of reusable expert cyber knowledge. The approach determines the cyber threat exposure of system assets (people, processes, and technologies), identifies the importance of each asset to mission success, and produces a prioritized list of recommended mitigations based on a stated level of risk tolerance. This paper briefly introduces the BluGen capability-based methodology, setting the stage for description of an application of BluGen to a notional space system and comparison to event-based risk assessment.

The remainder of this paper is organized as follows. First, we survey related work in the cyber risk assessment domain and relevant research about expert scoring. Then we summarize the BluGen capability-based risk assessment approach. Next, we

HoTSoS '18, April 10–11, 2018, Raleigh, NC, USA

ACM ISBN 978-1-4503-6455-3/18/04…$15.00
https://doi.org/10.1145/3190619.3190644

illustrate the risk assessment approach in the context of a space system. Finally, we compare the automated capability-based risk assessment approach versus the commonly used event-based approach, propose future work, and provide a summary.

We selected a space system as a compelling example to illustrate BluGen because space systems are increasingly dependent on cyber components and cyberattack is recognized as a significant threat [10]–[13] with several incidents publicly disclosed. Ground-based command and control systems for space systems are critical to telemetry, tracking, and commanding of space systems and their payloads. Securing these ground-based systems against cyberattack is a high stakes activity due to the cost of the space and ground segments and their importance to space mission success.

2 RELATED WORK

Cyber risk assessment. Several cyber risk assessment methodologies are described in the literature and in use today. These include Carnegie Mellon Software Engineering Institute Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) Allegro [14], ISACA Risk IT Framework based on Control Objectives for Information and Related Technologies (COBIT) [15], [16], MITRE Crown Jewels Analysis (CJA) and Threat Assessment and Remediation Methodology (TARA) [17], [18], US National Institute of Standards and Technology (NIST) Special Publication 800-30: Guide for Conducting Risk Assessments [19], and Johns Hopkins University Applied Physics Laboratory Mission Information Risk Analysis (MIRA) [20].

A common theme in these approaches is event-centricity. They all analyze risk by attempting to enumerate potential threatening events. While these methods do not require enumeration of all the events, oftentimes the quality of the assessment depends on how fully one can ensure that the list of enumerated events represents the actual threat environment of the system under analysis. The enumeration is then scored by experts in various qualitative ways to account for factors such as mission impact, likelihood of occurrence, level of adversary effort (LOE) required to realize the event, and sophistication of the threat actor. Below we discuss challenges inherent in event-based approaches, such as issues of repeatability, reproducibility, scalability.

In addition to event-based and capability-based risk assessment, two other classes of cyber risk assessments are compliance-based and loss-based. Compliance-based risk assessments aid organizations to comply with externally-imposed cybersecurity policies, such as the Payment Card Industry Data Security Standard (PCI DSS) [21]. Loss-based methodologies consider risk in terms of annualized loss expectancy (ALE) which is the product of single loss expectancy (SLE) and annual rate of occurrence (ARO) [22]. A discussion of these methodologies is beyond the scope of this paper. They are mentioned here for completeness.

Event-centric approaches typically rely on manually-derived, qualitative judgments by subject matter experts (SME) to evaluate threats, assign mission impacts, and assess likelihood of occurrence or level of effort required to realize each threat. The number of judgments required increases as the systems under assessment increase in complexity. For example, depending upon which event-based methodology is chosen, SMEs may have to produce up to 4,500 separate scores to assess a system with 3 missions, 25 assets, 4 data types per asset, 5 identified attack vectors and the traditional confidentiality/ integrity/ availability (C/I/A) cyber effects. In terms of time and attention required of SMEs doing the scoring, this is a scalability concern, particularly since most real-world systems are much larger and have more potential attack vectors. Consequently, for most of these types of assessments, SMEs only have time to consider a subset of the attack surface of the system, and thus may overlook significant attack opportunities. Moreover, the subset considered is often drawn according to expert judgment, and may not generalize to full system coverage, though they may assess the most critical assets in detail. In addition to scalability and coverage, the high demand for SME time is relevant from a skills availability standpoint. The information security non-profit ISACA forecasts "a global shortage of two million cyber security professionals by 2019." [23]

SME-based scoring is subject to human variability resulting in concerns about repeatability and reproducibility of assessment results. A measurement is repeatable if the same results are obtained when repeated by the same SME using the same methods and instruments. A measurement is reproducible if the same results are obtained by a different analyst or using different instruments. [24]–[26]. Due to human variability, it is possible that successive risk assessments of the same system may not yield consistent results.

Expert judgment. The term inter-rater reliability is used to describe the degree to which "different raters/observers give consistent estimates of the same phenomenon." [27] According to Trochim and Donnelly, "Whenever you use humans as a part of your measurement procedure, you have to worry about whether the results you get are reliable or consistent. People are notorious for their inconsistency. We are easily distractible. We get tired of doing repetitive tasks. We daydream. We misinterpret." This is a fair concern for SME-based risk assessment.

Holm et al. [28] justify the use of expert judgment in the context of decision support for the cybersecurity domain based on lack of a sufficient body of observations for the variables of interest. However, they point out that "when expert judgment is used, data quality is uncertain" due to errors in calibration (differences between judgments and ground truth).

A study by Hallberg et al. [29] explored inter-rater reliability and rater consensus when scoring the likelihood and severity of cyber security incidents. They found that "ratings of probability and severity are not reliable enough between raters to be considered a sound basis for the quantification of information security risks." They also found generally low consensus values and that experts and non-experts had similar degrees of consensus when performing ratings of cybersecurity incidents. Further, they observed that "low consensus values are not caused

by a few exceptionally hard incidents or a few poor raters but rather the rating being difficult in general."

Bolger and Wright [30] examined issues pertinent to the quality of expert judgment finding that the quality of the result depends on two main factors, the learnability of the domain in which judgment is applied and the degree to which the expert is experienced in making the type of judgment requested. Absent either of these factors, the authors assert that expert performance will be low. Learnability includes (a) availability of accurate, pertinent data or models on which to base judgments, (b) ability to express the judgments so that they can be verified, and (c) rapid, usable feedback which allows experts to refine their models and deepen their expertise. Of these, feedback is critically important when the domain evolves over time. Regarding SME-based cyber risk assessment, this study suggests that even SMEs who are experienced in the field may be disadvantaged due to lack of data, models, and feedback.

In current cyber risk assessment practice, expert scoring is often the instrument used to measure inputs, such as level of adversary effort, likelihood of adverse events, or importance of cyber components to mission success. Expert scoring is used because more evidence-based and deterministic tools do not yet exist for performing these measurements. We do not advocate or expect to eliminate SME judgment from risk assessment. In fact, most engineering disciplines rely on trained experts. Rather, we seek to reduce the volume of SME judgment required for individual assessments and redirect the energy of SMEs towards making their knowledge more accessible and reusable.

3 CAPABILITY-BASED RISK ASSESSMENT

This paper highlights a capability-based alternative to traditional cyber risk assessment embodied in a tool that we call BluGen. The major concepts of the BluGen approach are described here and illustrated in Figure 1.

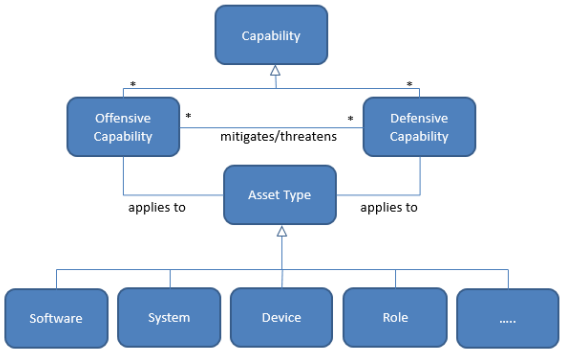

Figure 1. Reference Catalog Concepts

Instead of enumerating vulnerabilities, adverse events, and/or potential attack paths through a system, the capability-based approach focuses on describing the offensive capabilities possessed by an attacker (e.g. can exploit known vulnerabilities in software) and the opposing defensive capabilities that mitigate them (e.g. active patch management, hash-based white listing). Offensive capabilities are the tools and techniques that an adversary would use in a cyberattack. BluGen makes the following assumption: as the identified offensive capabilities are increasingly mitigated by defensive capabilities, the adversary's identified arsenal of effective offensive capabilities and the attacks he can compose are correspondingly reduced.

Asset types are the targets of both offensive and defensive capabilities. Defensive capabilities that mitigate offensive capabilities are represented in many-to-many "mitigates" mappings, which include effectiveness scores expressed as decimals, range 0.0 to 1.0. Conversely, offensive capabilities that threaten defensive capabilities are represented in many-to-many "threatens" mappings. Both offensive and defensive capabilities are mapped many-to-many to relevant asset types, as not all capabilities apply to all asset types.

Note that the mappings of capabilities to asset types and of defensive to offensive capabilities do not depend on the particulars of the system being analyzed. We believe that, once captured, this knowledge may be reusable for many systems. BluGen calls this reusable repository the Reference Catalog. Figure 2 provides an architectural overview of BluGen.

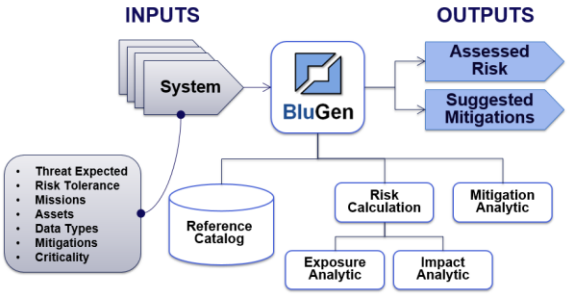

Figure 2. BluGen Architecture

Input Data Required by BluGen. The Reference Catalog is a key input necessary to perform a capability-centric risk assessment with BluGen. In addition, a description of the cyber system to be analyzed must be provided. This description includes the anticipated threat level, stakeholder-determined risk tolerance, entities in the system (instances of missions, assets, and data types), existing mitigations, and raw criticality scores. The inputs described thus far are similar to those required for event-based analysis, except for raw criticality scores, which we discuss in more detail below. In addition, each system asset must be associated with an asset type from the Reference Catalog. Also, a weight between 0.0 and 1.0 must be assigned to each mission, where the sum of all mission weights must equal 1.0. The anticipated threat level is a characterization of the sophistication of the attacker in terms of offensive capabilities that the system is expected to face. Risk tolerance is expressed in terms of maximum tolerable exposure (discussed below) and criticality, both expressed as values between 0.0 and 1.0 inclusive.

A raw criticality score, range 0.0 to 1.0, must be provided for every viable combination of the 4-tuple (M, A, D, E), where M represents a mission that the system supports, A represents a system asset that supports M, D represents a data type processed on A in support of M, and E represents a cyber effect (i.e. a breach of confidentiality, integrity, or availability). A viable combination is one where A processes D in support of M in the system being analyzed. A raw criticality score (R) of 0.0 means that the effect, E, on data, D, at asset, A, has no impact on the mission, M; whereas a score of 1.0 indicates total mission failure, with intermediate impact values possible in between these extremes. BluGen does not prescribe a method for determining the raw criticalities. They may, for example, be SME-generated (as is the case in the space example), or they may be determined in more evidence-based ways, such as from system sensors or by simulation.

Analysis Performed by BluGen. BluGen computes the risk associated with an asset as the combination of its exposure and criticality. BluGen defines criticality as the relative importance of each asset to the success of the mission(s) supported by the system being analyzed. An asset is more mission-critical if a greater number of highly-weighted missions rely on the asset and a greater number of important data types are processed there. The set of raw criticalities can be represented as 5-tuples (M, A, D, E, R). There may be up to three such tuples for each combination of M, A, and D, accounting for the three cyber effects (E). BluGen calculates the relative criticality of each asset as a weighted aggregation of the raw criticality scores provided as input during system characterization as follows.

First the raw criticality scores are simplified to 4-tuples (M, A, D, RH) where RH is the high-water mark of the three effect scores for a given M, A, and D. Next, BluGen calculates a weighted criticality for each asset from these tuples, multiplying RH times the mission weight for each present combination of M and A, then summing the mission-weighted scores for the asset. Finally, each of the mission-weighted asset criticalities is divided by the maximum asset criticality in order to scale the resulting values between 0.0 and 1.0. This calculation is illustrated in Figure 3.

```
max_asset_criticality = -1.0
For each A in the set of (M, A, D, R_H)
    criticality_A = 0
    For each M
        criticality_A = criticality_A + (R_H * weight_M)
    if (criticality_A > max_asset_criticality)
        max_asset_criticality = criticality_A
For each A in the set of (A, criticality_A)
    relative_ criticality_A = criticality_A / max_asset_criticality
```

Figure 3. Criticality Score Calculation

Exposure is the degree to which an asset is potentially threatened by unmitigated offensive capabilities believed to be possessed by the anticipated adversary. BluGen calculates the exposure of each asset in the cyber system based on its existing mitigations and the offensive capabilities that typically threaten an asset of its type as defined in the Reference Catalog. Figure 4 illustrates the exposure calculation for a notional system asset. An asset has a higher exposure to anticipated cyber threat actors if it is threatened by a greater number of offensive capabilities for which there are no corresponding defensive solutions.

BluGen first refers to the Reference Catalog for the asset under consideration to determine which offensive capabilities threaten assets of this type. In the example, the search yields offensive capabilities, OC1, OC2, and OC3. The Reference Catalog mappings also indicate the defensive solutions (collection of related defensive capabilities) that mitigate these threats with associated effectiveness scores, range 0.0 to 1.0, where 1.0 equates to 100% effective. For example, the defensive solution, DS1, mitigates threats to OC1 with 0.7 effectiveness. In the Reference Catalog, DS1 is composed of three defensive capabilities, DC1, DC2, and DC3, whose relative contributions to the solution are 0.6, 0.2, and 0.2 respectively. Contributions range from 0.0 to 1.0 and must sum to 1.0 within a solution. DC1 is designated as a required capability for the solution, while DC2 and DC3 are designated as optional. The system model reflects that the asset under consideration is already protected by DC1 and DC3, but DC2 is not present. To get any coverage credit for a given solution, the asset must have at least the required capabilities. If it does, the asset's coverage relative to the solution and threat is computed by multiplying the solution effectiveness times the relative contribution of each mitigation that is present and summing these products. If it does not have the required capabilities, the asset's coverage relative to the solution drops to zero.

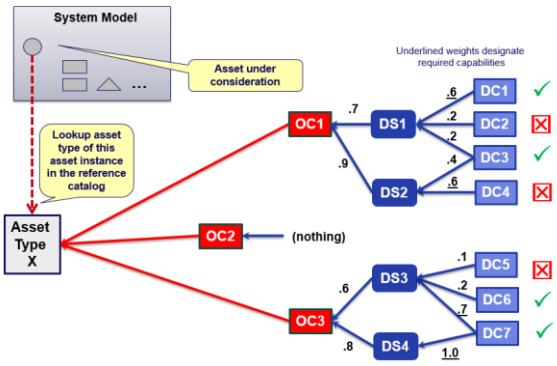

Figure 4. Exposure Score Example

In Figure 4, the asset has the required mitigation DC1; hence, its coverage relative to OC1 and DS1 is 0.7 times 0.6 (for DC1) plus 0.7 times 0.0 (for DC2 which is not present) plus 0.7 times 0.2 (for DC3) which equals 0.56. The coverage calculations for the remainder of the offensive capability/defensive solution pairs in the example are shown in Table 1.

To determine the asset's coverage, the calculation described in the preceding paragraph is repeated for each selected offensive capability and relevant defensive solution. For each offensive capability, BluGen chooses the "best" defensive solution, where best is defined as the solution that provides the highest coverage, while taking advantage of existing mitigations where possible. In this example, DS1 provides the best coverage

of the asset against OC1 at 0.56 and DS4 provides the best coverage against OC3 at 0.8. There is no solution for OC2; hence the asset's coverage relative to OC2 is 0. The asset's coverage values for each offensive capability and selected solution are summed, then this result is divided by the number of capabilities threatening the asset to yield an average coverage value for the asset. Exposure is equal to 1.0 minus coverage. As shown at the bottom of Table 1, the asset's average coverage is 0.45 and its exposure is 0.55.

Table 1. Exposure Score Example

			Contrib.	Exists	Coverage
OC1	DS1				
	0.7	DC1	0.6	Y	0.42
		DC2	0.2	N	0.0
		DC3	0.2	Y	0.14
					0.56
OC1	DS2				
	0.9	DC3	0.4	Y	0.36
		DC4	0.6	N	0.0
					0.0
OC2	n/a				0.0
OC3	DS3				
	0.6	DC5	0.1	N	0.0
		DC6	0.2	Y	0.12
		DC7	0.7	Y	0.42
					0.54
OC3	DS4				
	0.8	DC7	1.0	Y	0.80
	OC1	OC2	OC3	Sum	Score
Coverage	0.56	0	0.8	1.36	**0.45**
Exposure	0.44	1	0.2	1.64	**0.55**

To visualize the risk, BluGen plots assets on a scatter plot with exposure on the Y axis and criticality on the X axis as shown in the illustrative example in Section 4 Figure 8. Assets in the upper right quadrant of the plot (the red shaded area) are the assets whose exposure and criticality exceed the risk tolerance specified by the stakeholder.

Asset	Asset Type	Threat Capability	Mitigations	Mitigation Required?
Ground Segment Network Switch	Network Device	Find and Exploit Unknown Vulnerabilities in OS, firmware or application software on computing devices	Protect against unknown vulnerabilities (CVE and CWE) except in boot process	YES
			Limit impact of unknown vulnerabilities (CVE and CWE) on computing devices	NO
			Limit impact of unknown vulnerabilities (CVE and CWE) on computing devices without hypervisors	NO
			Detect exploitation on unknown vulnerabilities on computing devices	NO

Figure 5. Notional Reference Catalog Mapping

BluGen can also recommend mitigations to bring the exposures of highly critical assets into an acceptable range based on the specified risk tolerance. Figure 5 presents a notional illustration of the mitigation mapping for a single asset and threat. Starting from an asset's type, the offensive capabilities which threaten it are determined from information in the Reference Catalog. Likewise, defensive capabilities that mitigate the specific offensive capabilities are determined from the mappings in the Reference Catalog.

In addition to recommending mitigations to consider adding, BluGen can highlight existing mitigations that do not appear to be contributing to defense of a given asset based on the offensive capabilities that are known to threaten it.

4 APPLICATION OF BLUGEN TO A SPACE EXAMPLE

To illustrate the capability-based risk assessment, we consider a notional but representative space system that includes a ground segment, ground entry point (GEP) segment, and a space segment consisting of a satellite space vehicle. This system has been previously assessed for cyber risk using an event-based approach [31] and has also been studied for survivability [32] and resilience [33]. The system characterized here is representative of actual space missions, but is hypothetical and unclassified to enable sharing.

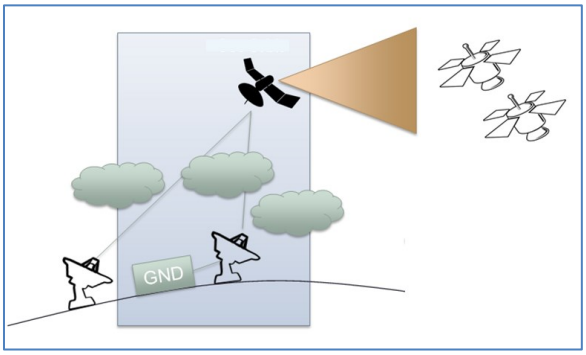

Figure 6. Space Situational Awareness System

The exemplar system, shown in Figure 6, facilitates Defensive Space Control (DSC), specifically space situational awareness (SSA). The system has two sub-missions supporting SSA: a communications mission and a sensing mission. In this example, we focus on the Ground Control Segment portion of the overall system. The ground segment, shown in Figure 7, has only basic cyber mitigations. We begin by collecting the data necessary to perform the capability-based assessment. We determine the entities in the ground segment: the missions, assets, and data types. The ground segment has 33 assets; we map each asset to a corresponding asset type in the Reference Catalog. For example, Data Switch 1 is a Network Device, Sensor Manager is a General User, GEP Crypto is an Endpoint Cryptographic Device, Storage Server is a Computing Device, and Storage Server Link is a Wired Link.

The ground segment has 25 data types and two missions, communications and sensing, which, according to SME judgment, we weight at 0.6 and 0.4 respectively. Assigning weights to the missions provides a means to indicate their relative importance. System characterization data can be gathered from drawings, documentation, network scans, experts, and other automated and manual methods. For the ground segment system, most of the necessary information, except for the mission weights and mapping of asset types to the Reference Catalog, was collected when the prior event-based assessment was performed in [31] and reused for the capability-based assessment.

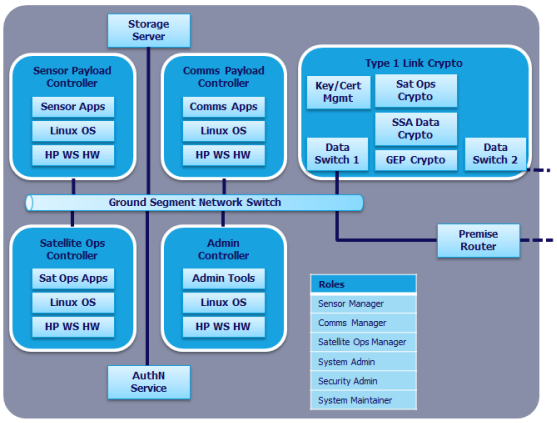

Figure 7. Ground Control Segment Architecture

Next, we define the relationships between the assets in the system. In event-based risk assessment, we are interested in connectivity between assets, but in BluGen we are primarily interested in which assets contain other assets as well as which assets inherit defenses from mitigations that reside on other assets. There are 32 containment relationships and 80 inheritance relationships in the ground segment. For example, a number of assets in the system inherit mitigation capabilities from the Authentication Service. Containment relationships are illustrated in Figure 7.

We also identify the existing mitigations or cyber defensive capabilities, if any, and the assets to which they apply. There are 204 existing defensive capabilities applied across 33 assets. In the notional example, Storage Server has defensive capabilities including Protect Against Known Vulnerabilities, and Detect and Respond to Malicious SW; while Authentication Service has defensive capabilities including Authenticate All Accounts, Detect and Respond to Authentication Attacks, and Prevent Use of Weak Passwords.

Finally, we provide raw criticality scores for each viable combination of mission (2), asset (33), data type (25), and cyber effect (3). An upper bound of possible scores is 4,950. Of these, 849 viable combinations were scored for mission criticality. In this example assessment, it was determined that, for the sensing mission, a confidentiality breach of Authentication Data on the Authentication Server would result in total mission failure (score = 1.0) while an integrity breach of Telemetry and Command Archive Logs on the Storage Server would leave the mission partially capable (score = 0.4). These are SME-generated inputs which must be provided by someone who understands the importance of the system assets and data processed on those assets to the missions supported by the system.

To perform the risk assessment, we must also indicate the anticipated threat level and risk tolerance as stated by the system stakeholder. For our example, we characterize the anticipated threat level as Advanced, chosen from possible values of Nascent, Limited, Moderate, and Advanced, and the maximum acceptable criticality and exposure at 0.5 and 0.25 respectively. BluGen calculates the criticality of each asset as an aggregation of the 849 raw criticality scores provided as input during system characterization. Furthermore, BluGen calculates the exposure of each asset taking into account the offensive-capability-to-asset-type and defensive-to-offensive-capability mappings in the Reference Catalog as well as the 204 existing mitigations.

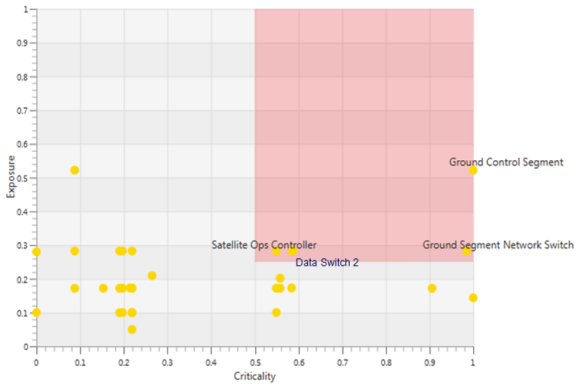

Figure 8. Risk Plot

Figure 8 shows the resulting risk plot. Four assets, Ground Control Segment, Ground Segment Network Switch, Satellite Ops Controller, and Data Link 2, are in the area of unacceptable risk.

The Ground Control Segment is an aggregated asset which contains all the other assets. Its overall exposure to the offensive cyber capabilities considered is 0.52, which exceeds the risk tolerance specified and indicates that additional mitigations are warranted. Each at 0.28 exposure, the Ground Segment Network Switch, Data Switch 2, and Satellite Ops Controller also exceed the risk tolerance specified and are key contributors to the overall exposure of the system. At 0.985 criticality, the Ground Segment Network Switch is identified as the most critical component of the system.

For the anticipated threat level, a capability-based analysis using BluGen identified a total of 835 unmitigated threats in the ground segment. By considering the recommendations in the mitigation report, the system security engineer (SSE) can apply additional mitigations with the objective of bringing the risk of all assets to an acceptable risk posture. BluGen recommends

mitigations for non-aggregate assets. It is up to the SSE to decide which mitigations to add and the best placement for the mitigations added. For example, since a common set of mitigations have been recommended for the Ground Segment Network Switch, Data Switch 2, and Satellite Ops Controller, the SSE may consider applying these mitigations on an asset from which the 3 assets can inherit capabilities. BluGen also identifies existing mitigations that may not be contributing to the protection of each non-aggregate asset. Again, it is up to the SSE to consider these recommendations holistically and decide upon the preferred mitigation strategy.

One way to achieve an acceptable risk posture is by addressing all of the 835 mitigation gaps identified, fully covering each asset to the maximum extent identified. The issue of how much coverage is enough is a stakeholder judgment captured in BluGen by specifying the risk tolerance. The commonly accepted strategies that organizations use to cope with risk are reduction, avoidance, transference, and acceptance [22]. Applying mitigations to bring the risk to an acceptable level is a form of risk reduction and may only partially cover the anticipated threats. Once the risk has been reduced to the desired tolerance, most organizations apply some combination of the other three approaches to the residual threat. For example, they may insure against some threats, if appropriate.

By using what-if analysis, it may be possible to achieve acceptable risk by applying fewer mitigations. BluGen makes it quick and easy for the SSE to try various possible combinations of mitigations and immediately see the impacts on risk. For example, Figure 9 shows a plot where, using half as many additional mitigations (411), an acceptable risk posture is achieved while at least partially covering each asset for each identified threat capability.

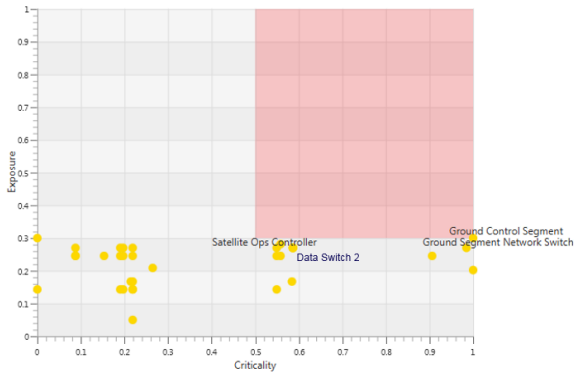

Figure 9. Risk Plot – Acceptable Coverage

5 COMPARISON

Both event-based and capability-based risk assessment methodologies follow these general steps: (1) characterize system, (2) characterize mission, (3) characterize threat, (4) assess risk, (5) determine mitigations, (6) apply mitigations and (7) reassess. As discussed earlier, at a conceptual level, the former grapples with risk by enumerating possible adverse events that a system may face; whereas, the latter considers risk based on presumed capabilities of the threat actor. Both approaches require knowledge of the assets, missions, data types, and mitigations that are present in the system.

Event-based analyses, particularly those approaches that utilize attack paths, require detailed information about the ways that system assets are interconnected. BluGen takes the most conservative stance on connectivity by assuming that every asset is reachable from every other asset in the system. This decision simplifies data collection and, more importantly, recognizes the fact that, in complex, highly interconnected systems, more connections may exist than are anticipated, particularly in systems that have been upgraded and evolved over time. For example, a 2013 data breach on a retailer's point of sale (POS) system was orchestrated, in part, due to a connection between the POS and a portal used by the retailer's refrigeration vendor, systems that one may not expect to be connected [34]. The capability-based approach requires information about containment (which assets contain other assets) and inheritance (which assets provide capabilities that mitigate threats faced by other assets). System characterization data can be challenging and labor-intensive to determine for either methodology. This information can be gathered from many sources, including drawings, documentation, network scans, experts, and other manual and automated methods (e.g. [35]).

Some event-based analysis methods require that two sets of scores be developed by SMEs to conduct the assessment: mission impact and level of effort or likelihood of occurrence. Mission impact scores are required for each viable combination of mission, asset, data type, and effect, while effort/likelihood scores are required for each viable combination of asset, data type, effect, and attack vector. On the other hand, to the extent that pre-existing Reference Catalog mappings can be reused, the BluGen capability-based analysis requires only one new set of scores, criticality scores for each viable combination of mission, asset, data type, and effect. It is often the case that scores must be SME-generated; therefore, fewer scores equates to significant time savings.

An event-based analysis of a similar Ground Segment was performed before the development of BluGen. Because this was a manual analysis, 22 threat capabilities were used to develop 60 representative attacks on the Ground Segment. Estimates were made for the adversary level of effort and impact of each attack. With only the selected attack sequences and estimated risk scores as background, the SSE proposed mitigations and rescored the attacks multiple times until an acceptable level of risk was achieved. By comparison, in one pass, BluGen automatically analyzed 323 threat capabilities, scored the risk, and proposed mitigations for each asset.

Many tool-assisted risk assessment methodologies provide mechanisms to save and reuse system characterization data [9], [18], [20] which supports reassessment. This is important because systems, the missions they serve, and the threats they face all evolve over time; thus, risk assessment is not a one-time

task. More broadly, inherent in the BluGen capability-based methodology is an attempt to capture for reuse a corpus of critical SME cyber knowledge known as the Reference Catalog: the mappings of (1) threats to asset types and (2) offensive to defensive capabilities. As illustrated in Figure 10, SME efforts to build the Reference Catalog as a community-managed and peer-reviewed asset may eventually make this knowledge more broadly available for reuse when assessing many systems, current and future. It is anticipated that the Reference Catalog will grow in depth and breadth over time.

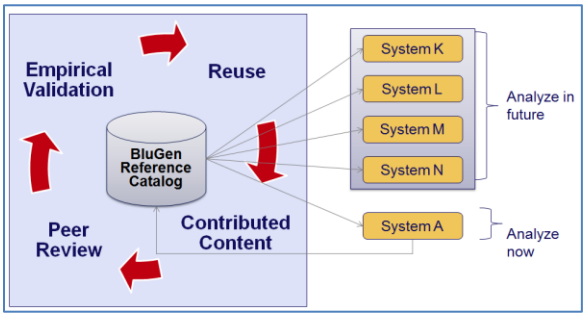

Figure 10. Reference Catalog Knowledge Reuse

BluGen outputs include a plot of assets based on their assessed criticality and exposure, as well as mitigation recommendations. A typical event-based assessment plots events, while mitigation determinations are usually not automated.

Finally, by reducing the number of scores that must be SME-generated and providing for substantial data and knowledge reuse, we believe the capability-based risk assessment embodied in BluGen takes steps towards making cyber risk assessments more repeatable and reproducible.

A validation experiment is underway to assess our intuition that BluGen improves coverage and efficiency of cyber risk assessment compared to manually-executed event-based assessments. Opportunities for rigorous evaluation of other aspects of BluGen are mentioned in the following section on future work.

6 LIMITATIONS AND FUTURE WORK

There are numerous opportunities to improve BluGen, and much work to be done to verify and validate the approach and tool. We briefly discuss some of them here, including expansion and verification of the Reference Catalog, streamlining data collection, developing additional analytics, and performing experiments to validate the analysis results.

Although significant work has been done on BluGen's Reference Catalog, this paper does not discuss the methods used to develop the catalog. In the future, a paper describing the approach and lessons-learned should be undertaken. Peer review of the details of the structure and content of the Reference Catalog would aid in validation.

Some may argue that the Reference Catalog represents an enormous SME footprint in its own right. This is certainly true in the current state of its development. Substantial manual SSE effort has gone into codifying the asset type-to-threat and threat-to-mitigation mappings thus far. Today the Reference Catalog is still in a relatively immature state, primarily encompassing the asset types, threats, and mitigations applicable to the space system analyzed herein. The applicability of the Reference Catalog to additional classes of systems will necessarily require the contribution of additional content to accommodate systems with asset types for which catalog data has not yet been captured. Once those mappings have been created, they may also be reused to assess other similar systems.

While SME input to the catalog is expected to diminish over time, the total elimination of SME input is not envisioned. Due to the dynamic nature of the cyber domain, there will continue to be new asset types, threats, and mitigations to be considered, requiring ongoing evolution of the Reference Catalog. SMEs will have a role in this evolution; however, there may be opportunities to support the SME efforts via automation. In addition, use of SMEs as peer reviewers of the catalog is a goal. We believe that creating a reusable knowledge base in the Reference Catalog is a better use of SME resources than using them as ephemeral measuring instruments.

Future emphasis on automated extraction of system characterization, either by direct analysis of the system itself or via ingest from system engineering drawings, such as from model-based system engineering (MBSE) tools, would significantly reduce the labor associated with describing the systems to be assessed. This is also the case for mission impact scoring [36], [37].

The development of additional analytics is an area that presents great opportunity. For example, cyber-enabled systems are made up of "hundreds of tradable variables that must be balanced in order to develop a viable system solution" (including cost, availability, feasibility, etc.) where each trade-off is a "compromise between objectives such that improving one requires that we degrade another" [38]. Algorithms which analyze the trade space of mitigation alternatives and offer balanced recommendations aligned with a specified set of stakeholder objectives would be valuable. Other examples of potential future analytics could encompass the time dimension of cyber threat and resilience considerations.

Finally, rigorous scientific experimentation is required to support or falsify certain claims about BluGen. For example, although we believe BluGen takes steps in the right direction toward making cyber risk assessments more repeatable and reproducible, additional work is needed to demonstrate it. In addition, empirical data should be analyzed to validate the mitigation-threat mappings and effectiveness scores in the Reference Catalog. Moreover, controlled experiments are needed to validate that results from BluGen are accurate representations of cyber risk for real-world systems as well as to raise confidence that mitigation recommendations result in improved security.

7 SUMMARY

This paper discussed the BluGen automated capability-based risk assessment methodology, comparing and contrasting it informally with the more common manual event-based methodology. We described the application of BluGen to a notional space system and explained how the results could be used by stakeholders and SSEs to assess cyber risk and make mitigation decisions for cyber systems. Although BluGen does not eliminate SMEs as a key element of cyber risk assessment, we believe that BluGen moves the SME to a better place in the risk assessment process, as a knowledge resource rather than a measuring instrument, potentially reducing the amount of new SME input that is required for performing cyber risk assessments.

ACKNOWLEDGEMENTS

The authors would like to thank the APL reviewers who contributed their thoughtful recommendations for the betterment of this paper. Additionally, we would like to acknowledge the resources and support that the Office of the Undersecretary of Defense (OUSD) for Acquisitions, Technology, and Logistics (AT&L) / Command, Control, Communications, Cyber, and Business Systems (C3CB) has provided in the development and maturation of BluGen.

REFERENCES

[1] Verizon, "2017 Data Breach Investigations Report," *Verizon Bus. J.*, no. 1, pp. 1–48, 2017.
[2] Microsoft, "Microsoft Security Intelligence Report," 2016.
[3] J. Gosler and L. Von Thaer, "Resilient Military Systems and the Advanced Cyber Threat," 2013.
[4] Ponemon Institute, "Cost of Cyber Crime Study: United States Benchmark Study of U.S. Companies," 2013.
[5] Center for Strategic and International Studies, "Net Losses: Estimating the Global Cost of Cybercrime," 2014.
[6] K. Chandrasekar *et al.*, "Symantec Internet Security Threat Report," 2017.
[7] Equifax, "Equifax Releases Details on Cybersecurity Incident, Announces Personnel Changes," 2017. [Online]. Available: https://www.equifaxsecurity2017.com/2017/09/15/equifax-releases-details-cybersecurity-incident-announces-personnel-changes/.
[8] Cybersecurity Ventures, "Cybersecurity Market Report," 2017. [Online]. Available: https://cybersecurityventures.com/cybersecurity-market-report/.
[9] T. Llanso, M. McNeil, D. Pearson, and G. Moore, "An Analytic Framework for Mission-Cyber Risk Assessment and Mitigation Recommendation," *Hawaii Int. Conf. Syst. Sci.*, p. 10, 2017.
[10] J. Stuart, "Comment: Satellite industry must invest in cyber security," *Financial Times*, 2015.
[11] P. Lewis, "Space Will Be the Next Frontier for Cyber Attackers," *WIRED*, 2016.
[12] J. Robinson, "Governance Challenges at the Intersection of Space and Cyber Security," *The Space Review*, 2016.
[13] N. Cohen, R. Ewart, W. Wheeler, and J. Betser, "Spacecraft Embedded Cyber Defense-Prototypes & Experimentation," *AIAA Space Forum*, 2016.
[14] R. A. Caralli, J. F. Stevens, L. R. Young, and W. R. Wilson, "Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process," 2007.
[15] ISACA, "The Risk IT Framework," 2009.
[16] R. A. M. Schmittling, "Performing a Security Risk Assessment," *ISACA J.*, vol. 1, 2010.
[17] J. Wynn *et al.*, "Threat Assessment & Remediation Analysis (TARA)," 2011.
[18] MITRE, "Cyber Mission Assurance Engineering: A Risk-Based, Threat-Informed Approach to Address Advanced Adversaries," 2013.
[19] National Institute of Standards and Technology, "National Institute of Standards and Technology Special Publication 800-30 R1: Guide for Conducting Risk Assessments," 2012.
[20] T. Llanso, P. A. Hamilton, and M. Silberglitt, "MAAP: Mission Assurance Analytics Platform," in *IEEE Conference on Technologies for Homeland Security (HST)*, 2012.
[21] PCI Security Standards Council, "PCI DSS Quick Reference Guide," 2015.
[22] D. Czagan, "Quantitative Risk Analysis," *INFOSEC Institute*, 2013. [Online]. Available: http://resources.infosecinstitute.com/quantitative-risk-analysis/.
[23] J. Kauflin, "The Fast-Growing Job With A Huge Skills Gap: Cyber Security," *Forbes*, 2017.
[24] J. Vitek and T. Kalibera, "Repeatability, reproducibility, and rigor in systems research," *Proc. ninth ACM Int. Conf. Embed. Softw. - EMSOFT '11*, p. 33, 2011.
[25] K. Bollen, J. Cacioppo, R. Kaplan, J. A. Krosnick, and J. L. Olds, "Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science," *Rep. Subcomm. Replicability Sci. Advis. Comm. to Natl. Sci. Found. Dir. Soc. Behav. Econ. Sci.*, pp. 1–29, 2015.
[26] B. Y. C. Collberg, T. A. Proebsting, W. H. En, C. Collberg, and T. A. Proebsting, "Repeatability in Computer Systems," *Commun. ACM*, vol. 59, no. 3, pp. 62–69, 2016.
[27] W. Trochim and J. Donnelly, *The Research Methods Knowledge Base*, 3rd ed. Atomic Dog Publishing Inc, 2006.
[28] H. Holm, T. Sommestad, M. Ekstedt, and N. Honeth, "Indicators of expert judgement and their significance: An empirical investigation in the area of cyber security," *Expert Syst.*, vol. 31, no. 4, pp. 299–318, 2014.
[29] J. Hallberg, J. Bengtsson, N. Hallberg, H. Karlzén, and T. Sommestad, "The Significance of Information Security Risk Assessments Exploring the Consensus of Raters' Perceptions of Probability and Severity," *Int. Conf. Secur. Manag.*, pp. 131–137, 2017.
[30] F. Bolger and G. Wright, "Assessing the Quality of Expert Judgment - Issues and Analysis," *Decis. Support Syst.*, vol. 11, no. 1, 1994.

[31] T. Llanso, G. Tally, M. Silberglitt, and T. Anderson, "Mission-Based Analysis For Assessing Cyber Risk In Critical Infrastructure Systems," *International Federation for Information Processing (IFIP) - Critical Infrastructure Protection VII*, vol. VII, J. Butts and S. Shenoi, Eds. Springer, 2013, pp. 135–148.
[32] C. Knez, T. Llanso, D. Pearson, T. Schonfeld, and K. Sotzen, "Lessons learned from applying cyber risk management and survivability concepts to a space mission," *IEEE Aerospace Conference Proceedings*, 2016, June, pp. 1–8.
[33] T. Llanso and D. Pearson, "Achieving Space Mission Resilience To Cyber Attack: Architectural Implications," *AAIA Space*, 2016.
[34] T. Radichel, "SANS Institute InfoSec Reading Room Case Study: Critical Controls that Could Have Prevented Target Breach," 2014.
[35] L. Buchanan, M. Larkin, and A. D'Amico, "Mission Assurance Proof-of-Concept: Mapping Dependencies among Cyber Assets, Missions, and Users," 2012.
[36] T. Llansó and E. Klatt, "CyMRisk: An approach for computing mission risk due to cyber attacks," *8th Annu. IEEE Int. Syst. Conf. SysCon 2014 - Proc.*, pp. 1–7, 2014.
[37] S. Musman, A. Temin, and M. Tanner, "Computing the Impact of Cyber Attacks on Complex Missions," *Proceedings of the 5th International Conference on Warfare and Security*, 2010, pp. 1–15.
[38] A. D. Maccalman, S. M. Sanchez, M. L. Mcdonald, A. T. Karl, and S. R. Goerger, "Tradespace Analysis for Multiple Performance Measures," *Proceedings of the 2016 Winter Simulation Conference*, 2016, pp. 3063–3074.