Program-Level Risk Orchestration in AI-Enabled Cyber Infrastructure: A Systems-of-Systems Management Approach

(Author Details)
Kumar Saurabh
ksaurabh.pm@gmail.com
PMI, USA

Abstract

Machine Learning (ML) and Deep Learning (DL) driven AI capabilities are being increasingly leveraged for detection, anomaly discovery, and autonomous response within, across, and through cyber infrastructures. Scaling such cyber defense capabilities, they are deployed less as standalone solutions but as programs consisting of autonomous subsystems, common data assets, and joint operational processes. However, scaling also creates systemic risks (e.g., cascading, latent, and indirect) and poses control risks (e.g., distributed, dual-use) that are not sufficiently recognized and managed under prevalent cybersecurity risk management (RM) models, which are generally limited to controls at the component level, local rather than program-wide risk visibility and mitigation. This paper posits a program-level risk orchestration framework and some of its design principles, which are based on systems-of-systems (SoS) engineering and established risk RM standards. Building on early work on AI-based intrusion detection systems, SoS engineering and interoperability, cybersecurity and cyber risk threat modeling, as well as formal risk governance and capital standards, the paper develops the thesis of conceptualizing and managing AI-enabled cyber infrastructure as a risk RM program of a special kind, namely a system-of-systems, in need of coordinated oversight and control. The RM program-level framework it proposes as a starting point combines technical (i.e., from AI subsystems) cyber risk signals with risk governance, investment, and economic decision models to structure the consistent and repeatable identification, assessment, prioritization, treatment, monitoring, and reporting of cyber risks across the program. To better support accountability and return on investment, the program-level management perspective further aligns the operations of the cybersecurity function with a standard RM lifecycle, investment criteria, and cost and benefit capital standards. The paper also positions the proposed framework with respect to recent and extant works, by offering a complementary management lens to a mostly algorithm-focused stream of research.

Keywords: AI-enabled cybersecurity, program-level risk orchestration, systems-of-systems engineering, cyber risk governance, intrusion detection systems, risk management standards.

DOI: 10.21590/ijhit.04.01-3.09

1. Introduction

The increasing complexity of today's cyber infrastructure has changed the nature of cybersecurity. On the one hand, the introduction of artificial intelligence (AI) to security monitoring, detection, and response systems has provided means to address more advanced threats. On the other hand, it has also led to new classes of risks and failures, going beyond traditional program and asset-centric cybersecurity risk management. Understanding why, how, and to what extent AI-enabled cybersecurity has changed over the years is therefore critical to the development of effective approaches to cyber risk governance in complex AI ecosystems.

1.1 Artificial Intelligence in Cybersecurity

Cybersecurity systems initially were built around relatively simplistic rule-based or signature-driven approaches to security analysis. Signature libraries were built on top of hard-coded, static thresholds and expert rules to build deterministic and interpretable detection behaviors. Although they were highly explainable, rule-based systems were fragile, slow to scale, and unable to detect novel attacks or adapt to new attack patterns or network dynamics (Wang et al., 2006).

In search of better ways to address evolving and large-scale threats, cybersecurity research began to incorporate data-driven and machine learning methods into security analytics. These approaches allowed systems to learn attack patterns from larger volumes of network or system data. Machine learning allowed for anomaly detection, behavioral profiling, and statistical classification of malicious activities (Wang et al., 2006).

Recent developments in deep learning further pushed the state of the art in intrusion and anomaly detection. Deep neural networks provided new means for hierarchical feature representation learning and generalizable anomaly and attack detection (Buczak & Guven, 2015; Ferrag et al., 2020).

As AI-based security capabilities matured, they have been applied not just to traditional enterprise or organizational network defense but also to a host of resource-constrained and distributed cyber infrastructures. Wireless sensor networks, IoT deployments, and some industrial networks are increasingly complex, heterogeneous, data-rich, decentralized, and resource-constrained settings. Thus, they are ill-suited to manually crafted or static security controls. AI-enabled intrusion and anomaly detection have therefore found widespread applications in these systems as well (Wang et al., 2006; Ferrag et al., 2020).

1.2 From Cybersecurity to Systemic Cyber Risk

AI has led to more robust and automated cyber defense mechanisms. Yet, it has also enabled the rise of systemic cyber risk. As part of more complex and automated enterprise cyber infrastructures, modern AI-enabled systems are now part of broader information, management, and operational architectures. These AI systems are often interdependent in terms of shared data

sources, training processes, common computing infrastructure, and even shared code, processes, or data artifacts. These interdependencies add up to create complex dependency structures across the enterprise and adjacent cyber systems.

AI security components' interdependence contributes to greater vulnerability to cascading failure. Faults or misclassifications on one component (such as a sensor network AI engine) can propagate downstream to multiple components. Inaccurate model outputs, bad data, or even undetected concept drifts can all lead to the simultaneous failure of multiple security functions in the absence of redundant controls (Sommer & Paxson, 2010). In this way, failures are no longer isolated to a single device or node as in a traditional infrastructure but can have system-wide implications due to automation and central orchestration (Weiss & Ngo, 2015).

Furthermore, AI systems add an additional layer of operational and management complexity. Security mechanisms built on AI are often adaptive, learning, and autonomous, complicating management, monitoring, or interventions using established methods. Actions in learning systems are no longer deterministic or known a priori, obscuring causality and accountability. In addition, many threat models often assume a rational and capable threat actor that can directly or indirectly manipulate the learning process (Schneier, 1999; Sommer & Paxson, 2010). In many cases, security functions or components influence each other through shared feedback loops or model training, giving rise to indirect and often opaque adaptive relationships.

AI-driven security capabilities thus produce cybersecurity risk that is no longer localized, independent, or resource-specific but becomes the product of multiple adaptive, interactive systems at scale.

1.3 Research Gap

While there is now a large body of literature focused on AI-based anomaly or intrusion detection, most of these efforts remain focused on understanding, benchmarking, or improving the performance of algorithms and models. The question of how these AI systems should be coordinated, managed, or governed within the larger program has not yet received significant attention in the literature (Berman et al., 2019).

On the other hand, research on systems-of-systems has a long history in engineering and has produced significant insight on how systems can be understood as interconnected but operationally independent, managerially autonomous, and emergent in behavior (Klein & Van Vliet, 2013). The study of systems-of-systems has also explored questions of how these systems can be governed and managed at the program level. However, the development of these principles is rarely linked back to concrete research and application in areas like AI cybersecurity.

Organizations are left without a theory of risk to guide the management of increasingly complex and interconnected cyber systems of systems. Much of current thinking on risk management still

focuses on risks at the system, node, or asset level and does not extend to the program and management levels (ISA, 2019). Existing approaches therefore often treat AI components as just another technical asset rather than a significant contributor to the overall program risk posture.

1.4 Research Objectives and Contributions

To help close this gap, this paper makes the following contributions and addresses three main objectives. First, it views AI-enabled cyber infrastructure from the lens of a system-of-systems, with interdependencies, emergent risk behaviors, and cross-program effects being explicitly modeled. Second, it proposes a framework for program-level risk orchestration that coordinates risk identification, analysis, and response across the AI of security system's different components. Third, it links cybersecurity operations and behavior to well-established risk governance and economic decision-making frameworks to enable more rational, transparent, and cost-effective security decisions.

By synthesizing AI-based cybersecurity research, systems-of-systems engineering, and formal risk management theory, this paper provides a more structured way for managing the growing class of cyber risk emerging from complex, AI-enabled cyber systems. This lens thus shifts attention from improving the performance of individual technical components to a more holistic understanding of overall cybersecurity resilience, governance, and long-term risk.

2. AI-Enabled Cyber Infrastructure as a System-of-Systems

AI-enabled cyber infrastructures are increasingly representative of a system-of-systems (SoS) where multiple constituent subsystems, though individually capable of performing autonomous functions, are coupled together to provide cybersecurity services. In modern cyber defense environments, artificial intelligence (AI) or machine learning (ML)-driven components are no longer restricted to a particular local environment, but span across multiple platforms, organizational units, and software/hardware layers. As a result, cyber infrastructures become distributed, modular, and adaptive. While this structural configuration empowers expansion and innovation at scale, it also opens doors to complex and emergent systemic risks which cannot be addressed with traditional component-focused strategies.

2.1 AI Cybersecurity Program Constituents

AI cybersecurity programs are typically comprised of multiple intelligent subsystems that are leveraged for their mutual security functions. These generally include subsystems with detection capabilities such as intrusion detection engines, anomaly detection engines, and other automated response mechanisms. Intrusion detection engines are AI algorithms which are trained to perform NIDS/NIPS or behavioral security analytics on network data, system logs, or application data. Such engines often learn from historical attack data and identify malicious activities using supervised or hybrid learning approaches (Apruzzese et al., 2018).

On the other hand, anomaly detection engines are detection algorithms which identify deviations in expected network or system behaviors which could indicate attacks in data where attack signatures are incomplete or missing. Due to high heterogeneity and irregular traffic patterns, IoT/wireless sensor networks are typical examples where such techniques are relevant (Zarpelão et al., 2017). In general, intrusion detection engines and anomaly detection engines make up the detection core of AI-enabled cybersecurity programs.

Automated response mechanisms, as the name suggests, are subsystems within these programs which carry out automated response or mitigation actions once they receive a signal or threat detection result from the detection subsystems. These response actions could range from simple alarms and alerts to automatic quarantine of a system or throttling of suspicious network traffic. The main caveat in the case of response mechanisms is that the integration and tight-coupling between them and the detection engines make the overall program more coupled, and incorrect detections could lead to misguided responses.

AI cybersecurity programs are also characterized by tight-coupling at the data, model, and infrastructure levels. It is common for training data, feature representations, and even trained models to be shared across multiple detection or response mechanisms in an AI-enabled cybersecurity program. The same also holds for deployment or runtime infrastructure, where these subsystems often run on the same server or hardware platform, cloud resources, or even distributed edge devices (Apruzzese et al., 2018; Zarpelão et al., 2017). The strong data and infrastructure sharing across components leads to high dependencies across the program, where changes or updates at one subsystem could potentially affect the entire program.

2.2 System-of-Systems Properties of AI Cybersecurity Programs

AI-enabled cyber infrastructures fit closely with the established systems-of-systems literature. First, the operational independence of the individual cybersecurity subsystems is a major defining property. This property refers to the subsystems' abilities to individually operate and still provide some form of standalone security value. In the case of the example discussed earlier, even if the anomaly detection engine is down, the intrusion detection engine could still be operational and monitor network data for attacks. The value of the overall cybersecurity program, however, is in the integration of their security functions, rather than the individual components (Maier, 1998).

Secondly, in the context of AI-enabled cyber infrastructures, individual subsystems are characterized by managerial independence and decentralized ownership. Different organizational groups, external partners, or vendors may be responsible for the development, deployment, or management of an AI cybersecurity component. Therefore, the authority over model changes, data sharing and usage policies, deployment and operational practices are not centralized (Dahmann & Baldwin, 2008). This disjoint management of multiple systems, naturally

complicates the attempts at centralized risk management and can lead to conflicting security objectives at an organizational level.

A third characteristic is the evolutionary nature of such systems. Components in AI-enabled cyber infrastructures tend to get updated independently at different points in time, e.g. via model retraining, algorithmic updates, or infrastructure improvements. This could include changes in their internal data collection or data access mechanisms, algorithmic modifications, or other operational parameter adjustments. Since these system elements are not coordinated and updated as a whole, the resulting system is never in a stable configuration and continuously changes with each constituent subsystem's updates. According to the systems-of-systems engineering literature, this results in emergent properties where the outcome or overall behavior of the system of systems cannot be known from the knowledge of the individual systems (Jamshidi, 2017). In the cybersecurity context, this could materialize in the form of unanticipated detection blind-spots, detection and response action loops, etc.

2.3 System-of-Systems Risk of AI-Enabled Cyber Infrastructure

AI-enabled cyber infrastructures, as an instance of system-of-systems, are subject to a distinct set of program risks. In particular, the first point of consideration is the non-linear propagation of cyber risk through such programs. Individual or localized issues, such as a subtle bias in training data or a latent performance degradation of a model, are not isolated in these systems and could manifest across the entire AI-enabled cybersecurity program through common datasets, model sharing, and infrastructure components. In practice, these cascading effects are often difficult to identify using conventional risk assessments which focus on individual components instead of the whole. Another key implication is the inherent conflict between local and global optimizations. As the previous discussion has highlighted, individual subsystems are largely optimized at a local level, with their global effect on the entire program's resilience being a secondary consideration. To further exemplify this point, an automation policy may be designed to focus on response speed, at the system level resulting in unnecessary service disruptions and operational inefficiencies. These examples highlight the futility of siloed risk assessment and the need for a holistic, AI cybersecurity program-level risk orchestration.

Table 1. Mapping Systems-of-Systems Characteristics to AI-Enabled Cyber Infrastructure

SoS Characteristic	Definition in SoS Theory	Manifestation in AI-Enabled Cyber Infrastructure	Risk Implications
Operational Independence	Subsystems operate independently	IDS, anomaly detectors, and response engines function autonomously	Local failures may go unnoticed until systemic impact occurs

Managerial Independence	Decentralized control and ownership	Multiple teams or vendors manage AI security components	Inconsistent risk priorities and governance gaps
Evolutionary Development	Asynchronous subsystem evolution	Continuous model retraining and infrastructure updates	Configuration drift and unpredictable interactions
Emergent Behavior	System-level outcomes not predictable from components	Combined AI decisions create unexpected security states	Difficulty in anticipating cascading cyber risks
Interdependence	Functional reliance among subsystems	Shared data, models, and platforms	Amplified risk propagation across programs

3. Cyber Risk Dynamics in AI-Driven Security Systems

AI-driven security systems can change the nature of cyber risk by introducing the ability to learn, adapt, and execute automated decisions within key security functions. These systems create potential risks of their own by coupling the behavior of security functions to data sources, learned representations, and decision models. They can also add new dimensions of coupling between programs that share data, computational resources, and control policies. This is in contrast to traditional security architectures, where many dimensions of risk can be well-characterized and analyzed independently. In AI-enabled systems, cyber risk is more dynamic and emergent, linked to the adaptive and interactive nature of the underlying networks and adversary.

3.1 Threat Modeling for Learning-Based Cyber Defense

Threat models for learning-based cyber defense extend existing considerations to include adversarial interaction with the learning process. Attackers may also target the data, model, or resulting decisions. Traditional anomaly-based IDS/IPS (intrusion detection system/prevention system) approaches build a model of normal activity by identifying common patterns in network traffic, system calls, or user behavior and then detecting malicious activity as anomalies relative to the learned model (Garcia-Teodoro et al., 2009). Network level anomalies can include unusual packet headers, volumes, traffic distribution, content, protocol misuse, TCP flow characteristics, TLS handshake information, and session activity. In learning-based systems, the definition of anomalies are defined by the data and model, rather than explicitly specified. These are also likely to be more sensitive to novel or shifting behavior as well as adversarial manipulation (Nippold et al., 2020).

Learning systems are more susceptible to data poisoning, evasion, and model manipulation. In data poisoning, an adversary injects false, misleading, or malicious samples into a training dataset, which corrupts the resulting learned representations and compromises detection. Because many learning-based security models are continuously or periodically retrained, the poisoned data can propagate effects widely. Evasion attacks craft inputs to the classifier that the system will identify as benign, while retaining their malicious functionality. These attacks work by identifying decision boundaries learned during the training process, which can be subverted through adversarial perturbation of the features (Goodfellow et al., 2015). Evasion attacks are statistical in nature, instead of targeting specific rules as is often the case in non-learning systems.

Attackers can also manipulate learning-based models if they are able to observe, control, or otherwise influence the training process, model parameters, or chosen features. Schneier (1999) notes that when building threat models, designers should assume that attackers are adaptive agents that will actively explore, probe, and counteract detection mechanisms. In learning-based systems, this may mean that the target of the attack becomes the model itself as a result of these interactions. Cyber risk in AI-driven security is thus no longer solely external but can include the systems of learning and inference themselves (Garcia-Teodoro et al., 2009).

3.2 Operational Limitations of AI-Based Intrusion Detection

The operational performance of AI-based intrusion detection systems is constrained by training data. Performance is highly dependent on the training data being representative and comprehensive with respect to both normal and malicious behavior. In practice, such data may be incomplete, unbalanced, or biased towards particular kinds of attacks. This limits the detection model's ability to generalize to new attack types or normal behavior outside of the training data (Butun et al., 2013). This dependence on data introduces systemic risk, whereby detection performance can degrade significantly if the operational conditions of the system differ from the training data assumptions.

Model drift, where the statistical properties of input data change over time due to changes in network traffic patterns, user behavior, and system configurations can also present a serious performance limitation for deep learning models. Drift can be introduced by benign factors such as evolving normal network behavior, new applications, or emerging threats. Over time, drift will degrade model accuracy, often without operator visibility. Ferrag et al. (2020) note that many deep learning-based network intrusion detection systems show very high accuracy rates in experimental settings, but often struggle to translate to real-world performance. Without appropriate drift detection and retraining processes, model performance can silently degrade over time, increasing the risk of undetected attack.

Deployment and operation of AI-based intrusion detection systems are themselves non-trivial, and present many additional risk factors. Effective intrusion detection in practical settings

requires well-configured, reliable data sources, computational infrastructure, and integration with broader security management workflows. Many application settings such as wireless sensor networks and IoT present challenging environments for AI-based detection due to resource constraints and highly heterogeneous platforms and devices (Butun et al., 2013). Limited interpretability of learned representations can also pose challenges for incident response and operator trust, creating more friction in analysis and greater dependence on human operators. As a result, these risk factors demonstrate that the adoption of AI-based intrusion detection redistributes, rather than removes cyber risk, across data, models, infrastructure, and operator factors.

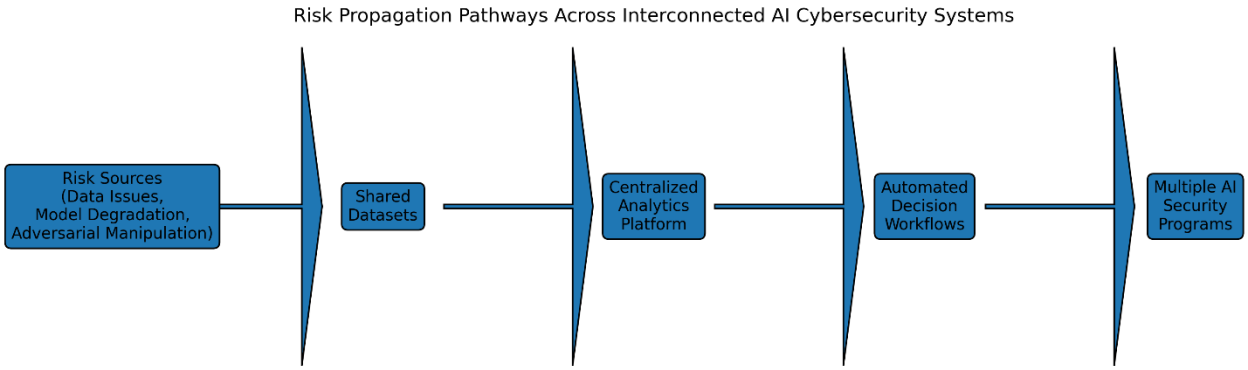
3.3 Program-Level Risk Amplification Mechanisms

Risk at the program or application level in AI-enabled cyber systems can be strongly coupled across applications through shared resources and centralization of certain activities. Risk can propagate across programs by sharing the same datasets as input to multiple detection models or IDS/IPS programs. While sharing of data may bring advantages of consistency and efficiency, if a common dataset is subject to contamination, bias, or other issues, multiple systems will be simultaneously affected. This can effectively transform what is otherwise a local data quality problem into a multi-system program-level vulnerability.

Coupling among security programs is often further amplified by centralization of analytics platforms, infrastructure, and operational policies. Shared training pipelines, feature extraction services, or detection and response decision engines create dependencies between security programs that can facilitate cascading failure. A fault, misconfiguration, or adversarial compromise at the platform level can have negative implications for detection and response capability across many independent programs. Localized model failures can also impact other automated response policies, threat intelligence sharing, or resource allocation decisions, increasing the broader effects.

In the absence of monitoring and escalation of failures, these can remain hidden until they surface as larger incidents. Figure 1 (Graph) provides a visual depiction of the typical risk propagation pathways within an interconnected system of AI-based cybersecurity applications. These show how issues in data quality, model performance, or adversary interaction may cascade to other systems through shared infrastructure and automated decision processes. These program and orchestration dynamics demonstrate the importance of managing program level risk orchestration to manage dependencies between AI cybersecurity systems.

Figure 1 (Graph): Risk Propagation Pathways Across Interconnected AI Cybersecurity Systems



4. Conceptual Foundations of Program-Level Risk Orchestration

Program-level risk orchestration is predicated on a recognition that contemporary AI-enabled cyber infrastructures manifest and operate as networked systems, not as siloed technical assets. As such, cyber risk management needs a framework that is more holistic, systematic, and cognizant of the economically complex, emergent, and cross-systemic nature of modern cyber risk. This section sets the conceptual stage for the idea of program-level risk orchestration, by contrasting its approach to traditional cyber risk frameworks, defining what is meant by risk orchestration as a management paradigm, and providing an economic foundation for program-level cyber risk decisions.

4.1 Limitations of Traditional Cyber Risk Management

Traditional cyber risk management frameworks are generally system-specific, centered on discrete technologies, applications, or business units. Classic NIST-style approaches, for example, focus on steps like asset identification, likelihood estimation, vulnerability analysis, and impact assessment (Stoneburner et al., 2002). Applied to the problem of managing risk in AI-driven cybersecurity tools, this “system-centric” (Stoneburner et al., 2002) risk framework leads to isolated assessments that are unaware of interdependencies and lack alignment.

Isolation and fragmentation are the direct result of each system being assessed in a vacuum, without accounting for differences in the underlying assumptions, measurement, and risk tolerance between systems. Controls are optimized for local effectiveness but lack an awareness of cross-system redundancies or gaps, mitigation actions are duplicated and not coordinated, and incident response remains myopic with respect to how risks propagate (Purdy, 2010). When AI-powered cyber tools share data sources, security models, and technical orchestration platforms, the feedback between local risk controls and system-level vulnerability profiles means that localized actions can actually escalate program risk.

The second key shortcoming is a lack of visibility and accountability for emergent, cross-systemic, and adaptive risks. Traditional risk assessment implies well-understood and demarcated systems (i.e., computing, AI models, operations) with consistent component behavior (e.g., likelihood and impact estimates) and threat models. In contrast, modern AI-enabled cyber infrastructures are characterized by complex interactions, adaptive learning, and feedback dynamics, which can produce system-level emergent behaviors and vulnerabilities that are not localized to a single component or team (Blank & Gallagher, 2012). A NIST-style, system-centric approach to risk assessment therefore has significant blind spots for these cascading, correlated, and economically complex risks.

4.2 Risk Orchestration as a Management Paradigm

Risk orchestration is a management process through which interdependencies between components are actively coordinated by aligning their information flows and decision rights. Applied to AI-enabled cyber risk, program-level risk orchestration focuses on cyber risk sensing, analysis, and response activities that are shared across related systems, rather than optimizing each individual risk register as a standalone entity. In this way, risk orchestration retains the enterprise risk management focus on material risks to the business, but differentiates itself by including an explicit emphasis on the dynamic interactions and emergent dynamics in networked systems.

A first critical distinction is between risk aggregation and risk orchestration. Risk aggregation generally involves a process of combining risk exposures from individual systems, such as building a program-level risk register, into a holistic view (Purdy, 2010). Risk aggregation can serve many important purposes, such as reporting, threat prioritization, communication, or intelligence fusion (Lalonde & Boiral, 2012). However, the aggregation process alone is generally insufficient to resolve differences in timing, semantics, measurements, and optimization strategies across disparate components. Risk orchestration, on the other hand, emphasizes alignment. Risk signals from individual AI security systems are mapped onto a common analytical base, calibrated to a common understanding of risk framing (likelihood and impact), and have associated escalation structures that support coordinated action (Lalonde & Boiral, 2012).

A second distinction is that risk orchestration is not simply a technical process but a governance imperative. Purdy (2010) argues that the inherently distributed nature of risk management in information security requires a coordination framework that spans both technical activities and strategic oversight. The rise of AI-enabled cybersecurity tools makes the link between distributed technical anomalies and strategic business risk even more tenuous (Abrams, 2022). Program-level risk orchestration therefore serves as a managerial construct that enables translation of technical data into cross-system risk intelligence that is both measurable and actionable.

4.3 Economic Perspectives on Cyber Risk Decisions

The need for a program-level risk orchestration process also has an economic explanation. Gordon and Loeb (2002) show that the economically optimal level of security investment in response to uncertainty is bounded and not directly proportional to perceived cyber risk. Spending more to avoid uncertain risk can be subject to severe diminishing returns, while spending less on shared risk-critical infrastructure can leave a program vulnerable to highly disproportional losses.

Component-level cyber risk investment decisions, if made in isolation, can often lead to economically inefficient outcomes. Redundant controls, misaligned priorities, and blind spots for systemic interdependencies are some common symptoms of a failure to align the investment decisions made by security and data owners (Anderson & Moore, 2006). At the same time, incentive misalignment in information security (Anderson & Moore, 2006) often leads to a situation where organizations reflexively overprotect their AI security assets but then underprotect other components which generate positive (or negative) externalities in shared security infrastructure.

Program-level risk orchestration makes the shift from optimizing component level security to system-wide economic efficiency. The investments made to secure individual AI security subsystems are guided not only by local risk assessments but also by an alignment to the program's aggregate risk posture and interdependency profile. In other words, the cross-system risk orchestration function can support a more economically efficient cybersecurity strategy, by calibrating security resilience to aggregated risk across the entire cyber program.

Table 2. Traditional Risk Management vs Program-Level Risk Orchestration

Dimension	**Traditional Risk Management**	**Program-Level Risk Orchestration**
Risk Scope	Individual systems or assets	Interconnected programs and systems
Assessment Approach	Siloed, system-centric	Coordinated, cross-system
Handling of Emergent Risk	Limited or implicit	Explicitly addressed
Decision Authority	Decentralized	Integrated and aligned
Investment Logic	Component-level optimization	Program-level efficiency
Governance Focus	Compliance-driven	Resilience and coordination-driven

5. Proposed Program-Level Risk Orchestration Framework

This section presents a structured program-level risk orchestration framework for AI-enabled cyber infrastructure, grounded in systems-of-systems engineering and standardized risk management principles. The framework addresses the limitations of isolated, component-level cybersecurity controls by enabling coordinated visibility, authority, and learning across interconnected AI security systems.

5.1 Design Principles

The proposed framework is guided by three foundational design principles derived from systems-of-systems theory and program management research.

System-wide visibility of AI security risks is essential due to the distributed and interdependent nature of AI-enabled cyber infrastructures. Individual intrusion detection systems, anomaly detectors, and response engines often operate with partial situational awareness, leading to fragmented risk assessments. Program-level orchestration establishes shared risk visibility by aggregating threat indicators, model performance metrics, and operational anomalies across subsystems. This holistic view enables the identification of emergent risks that would remain undetected within siloed security components. Such visibility is a core requirement in systems-of-systems environments, where system behavior cannot be inferred solely from individual elements (Maier, 1998).

Coordinated authority and escalation paths ensure that risk decisions align with program-level objectives rather than local optimization. In many AI security deployments, subsystems operate under separate ownership and governance structures, complicating response coordination during incidents. The proposed framework defines clear escalation mechanisms that link operational security signals to program-level decision authorities. This coordination enables timely prioritization of mitigation actions, allocation of resources, and policy adjustments across interconnected systems. Systems-of-systems engineering emphasizes the necessity of shared decision authority to manage emergent behavior and prevent fragmented control (Jamshidi, 2017).

Continuous monitoring and adaptive learning recognize that AI-enabled cyber risks evolve over time due to changes in threat behavior, data distributions, and system configurations. Static risk assessments are insufficient in environments characterized by learning-based security mechanisms. The framework integrates continuous monitoring of both technical indicators and risk outcomes, enabling feedback-driven adjustments to detection thresholds, response strategies, and governance policies. This adaptive capability supports long-term resilience by ensuring that risk management practices evolve alongside the systems they govern.

5.2 Structural Layers of the Framework

The program-level risk orchestration framework is structured into three interdependent layers that collectively enable coordinated risk governance.

The strategic governance and policy layer provides overarching direction and accountability for cyber risk management. This layer defines risk appetite, governance policies, compliance requirements, and investment priorities. It translates organizational objectives into actionable risk management mandates and ensures alignment between cybersecurity operations and executive decision-making. By operating at the program level, this layer addresses cross-system dependencies that are often overlooked in component-focused governance models.

The operational AI security layer encompasses the technical subsystems responsible for threat detection, anomaly identification, and automated response. These include machine learning and deep learning-based intrusion detection systems, monitoring tools, and response mechanisms deployed across the cyber infrastructure. While these systems retain operational independence, their outputs are structured to feed into program-level risk coordination processes.

The risk coordination and intelligence layer functions as the integrative core of the framework. It aggregates risk signals from operational systems, assesses their program-level implications, and facilitates communication between technical and strategic layers. This layer supports risk prioritization, escalation, and cross-system synchronization, ensuring that localized security events are evaluated in the context of overall program risk exposure.

5.3 Program-Level Risk Lifecycle

The framework operationalizes risk orchestration through a continuous lifecycle aligned with established risk management standards. Risk identification involves detecting threats, vulnerabilities, and anomalies across AI security subsystems. Risk prioritization evaluates identified risks based on potential impact, likelihood, and systemic propagation effects. Risk mitigation coordinates response actions across systems, including technical controls, policy adjustments, and resource reallocation. Risk monitoring tracks the effectiveness of mitigation measures and evolving threat conditions. Finally, feedback and learning integrate lessons from past incidents into future risk assessments and governance decisions, reinforcing adaptive risk management practices (Lalonde & Boiral, 2012).

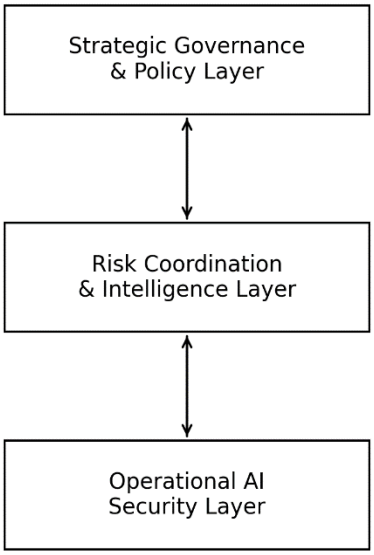
Table 3: Roles, Responsibilities, and Information Flows in Risk Orchestration

Role	Key Responsibilities	Information Exchanged
Strategic Governance Body	Define risk appetite and policies	Risk summaries, escalation reports
Program Risk Coordinator	Aggregate and prioritize risks	Threat intelligence, impact assessments

AI Security Operations	Detect and respond to threats	Alerts, performance metrics
Executive Decision Makers	Approve mitigation and investments	Risk trade-off analyses

Figure 2: Program-Level Risk Orchestration Architecture for AI-Enabled Cyber Infrastructure

The figure illustrates the layered interaction between governance, operational AI security systems, and the risk coordination layer, highlighting bidirectional information flows and escalation pathways that enable coordinated, program-level risk management.



6. Governance and Decision-Making Implications

6.1 Program-level Risk Orchestration

As the complexity of AI-enabled cyber infrastructure increases, new program level governance structures must be developed in order to facilitate accountability. In a cyber environment with multiple autonomous programs that house capabilities related to cybersecurity, the risks are no longer easily contained within a specific program. From the perspective of systems of systems, there may be a number of AI security-related subsystems that are fully independent from an operational and management standpoint, yet there may be behaviors across these systems that define cyber risk for the organization. Therefore, there needs to be a governance structure at the program level in order to provide effective cross-program cyber risk accountability.

Cyber risk ownership in an AI-enabled cyber infrastructure may need to be considered a cross-program responsibility. In other words, the cyber risk may not be the responsibility of one or more system owners, but rather it is the joint responsibility of a number of different programs,

platforms, and operational domains. The advantage of this type of cross-cutting ownership is that it can be used to determine dependencies that are invisible at the system level, such as a data pipeline, shared ML models, central response system, and so on. Without this type of ownership, one program may take action to reduce its risk, without understanding how this may affect the risk posture of other systems, which may increase the vulnerability of other systems, and lead to organizational impact at the infrastructure level.

In addition, the incorporation of technical risk indicators into governance decision-making is also important. This is the idea that AI-enabled cybersecurity systems produce large amounts of technical risk information, such as anomaly scores, detection confidence values, false positive rates, model error rates, and so on, which are important at the tactical and operational level. However, in some organizations, these types of data are not incorporated into decision-making as the governance structure or cadence for the organization may not have a formal structure in place to translate those indicators into governance-relevant decisions.

For a systems of systems governance perspective, this does not necessarily mean that an executive must understand all of the underlying mathematics of the algorithms in use. However, it does mean that the executive needs to have the understanding of how those technical risks impact other executive-focused concerns such as mission continuity, monetary exposure, legal or regulatory risk, compliance requirements, brand, and so on. Dahmann and Baldwin point out that governance in systems of systems environments must be able to achieve the right balance of decentralization along with a level of coordination oversight across different systems in the infrastructure. This is the argument for risk orchestration at the program level in AI-enabled cyber infrastructure, which translates risk information that has been captured from various components in order to be more aligned with governance-relevant information to feed into decision making by executives in the organization without diminishing the independence of different programs in the infrastructure.

6.2 Risk-informed Security Investment and Budgeting

Governing AI-enabled cyber infrastructure is also a problem of security investment and budgeting for AI-related capabilities. This can be different from traditional security environments where controls and risk indicators can be considered independent of each other. In an AI-enabled environment, as the number of AI-based cybersecurity systems and controls increases, the relationship between investment and risk may no longer be linear. In information security economics, when a certain investment threshold has been reached, increasing investment will provide a lower rate of return than prior investments. In other words, there is a point where throwing more money at a problem will not significantly increase protection (Gordon & Loeb, 2002).

Program-level risk orchestration can be used to identify the areas that will have the most value from cybersecurity investment. In some organizations, cybersecurity investment may be based

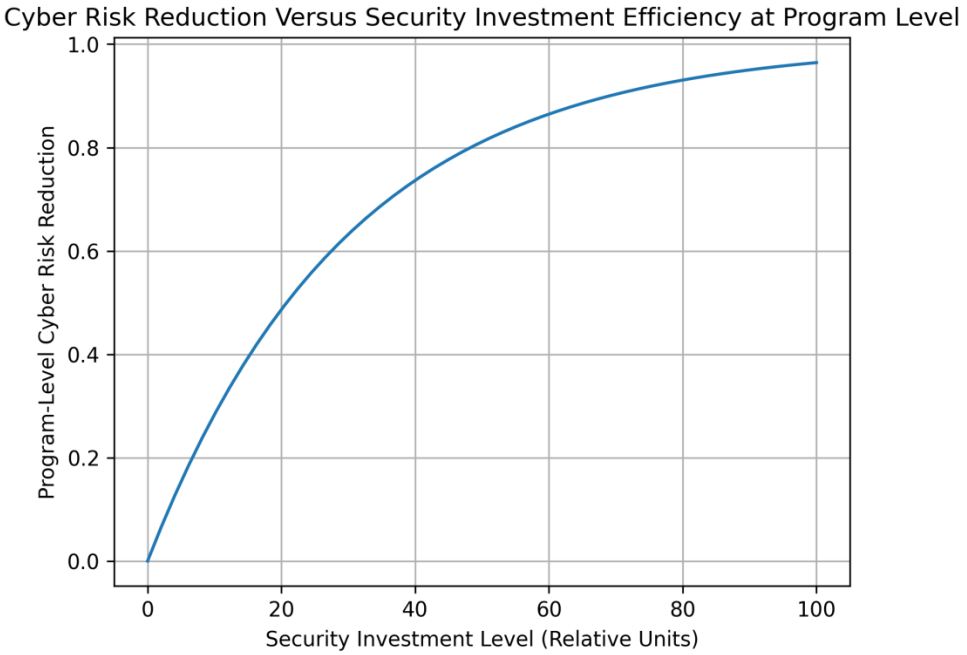
on the specific vulnerabilities in a single system. While there may be some high risk vulnerabilities in an individual system, this system may be less sensitive to those risks in terms of how the overall risk propagates across the different AI security programs in the organization. By taking a more crosscutting risk view, instead of considering component-specific vulnerabilities, it is possible to identify the risk trade-offs that can be addressed based on a security investment. Therefore, instead of uneven controls and protection that is only optimized for a specific program, it is possible to use a risk-informed approach to identifying cybersecurity investments that can have a higher value in terms of the infrastructure as a whole.

Program-level risk orchestration can also be used to avoid redundant or misaligned AI security investments. In an infrastructure with multiple autonomous cybersecurity programs, different programs may start investing in the same AI-based detection tools, analytics platforms, and monitoring and risk information systems. This may lead to an increased perception of security within an organization, but can also lead to unnecessary complexity, operational inefficiencies, and integration issues in some situations. In addition, if the systems are independently making assessments regarding risk, there may be conflicting risk estimates that can create accountability challenges and incident response delays.

In cybersecurity economics, investment in information security systems can be tied to incentives and the risk or threat environment in which the organization is operating (Anderson & Moore, 2006). In program-level orchestration, there are a number of components of risk that can be crosscutting for an organization. This can be used by decision makers to identify areas of investment or inefficiency based on how the cost relates to the risk or threat environment as well as organizational incentives related to risk management and response.

Figure 3. Cyber Risk Reduction Versus Security Investment Efficiency at Program Level

Figure 3 illustrates the relationship between cumulative cybersecurity investment and program-level risk reduction in AI-enabled cyber infrastructure. The graph demonstrates an initial phase of significant risk reduction as foundational controls are implemented, followed by a plateau where additional investment yields diminishing returns. The figure underscores the value of risk-informed orchestration in identifying optimal investment levels that balance security effectiveness with economic efficiency.



7. Discussion

7.1 Benefits of SoS–Based Risk Orchestration in AI-Enabled Systems

The first benefit of using an SoS perspective for risk orchestration is better alignment with the fundamental nature of cyber risk. Because an AI-enabled detection or response system may comprise multiple interacting security engines or subsystems, understanding the AI component of a particular system in isolation risks overlooking emergent properties of the system as a whole. That is, the optimal behavior of one subsystem (e.g., for detection accuracy or response time) may not contribute to program-level resilience (Jamshidi, 2017; Maier, 1998). In contrast, an SoS–based risk orchestration would require such an alignment of subsystem and program-level objectives, minimizing the risk of unforeseen interdependencies and failure cascades between otherwise independent cyber infrastructures.

Second, program-level orchestration could support better situational awareness in environments where an AI technique is used to help detect or respond to attacks. For example, rather than performing an independent and possibly suboptimal risk assessment, each security subsystem and operational environment could send a standardized signal to a coordinator or system-of-interest (SoI) representative, who would then correlate such signals and determine whether (and what type of) risk is present (Garcia-Teodoro et al., 2009; Sommer & Paxson, 2010). This type of orchestration would be particularly helpful in AI environments, where attacks may be

designed to have subversive signatures that are difficult to detect (i.e., distributed and more subtle than technical vulnerabilities) but are still interpretable by higher-level risk governance.

The third benefit of SoS–based orchestration of cybersecurity and AI-related risk is an improved understanding of roles, responsibilities, and authority with respect to program-level and enterprise risk (Lalonde & Boiral, 2012; Purdy, 2010). This follows naturally from SoS–based orchestration, in which responsibility and authority are delegated both vertically and horizontally in contrast to component-level approaches to cybersecurity (Stoneburner et al., 2002; Blank & Gallagher, 2012). In addition to the governance benefits of standardization, from a program-level perspective it also allows for a more flexible approach to cybersecurity, which is necessary given the dynamic and adaptive nature of AI-enabled systems.

7.2 Practical Considerations in Coordinating Security Operations across AI-Enabled Systems

While there are clear benefits to SoS–based orchestration of risk in AI-enabled systems, there are a number of practical challenges that must be considered. For example, because different subsystems may implement different AI or machine learning (ML) techniques, the variety of detection, analytics, or response systems in an SoS may preclude practical orchestration and coordination of cybersecurity and AI-related risk. This difficulty may be compounded by legacy systems and architectures, lack of standardization, and organizational silos and constraints that frequently characterize security operations, in particular with respect to intrusion detection and analysis (Butun et al., 2013; Zarpelão et al., 2017).

Another difficulty in coordinating and orchestrating risk in AI-enabled systems is the often opaque and thus difficult to explain ML and AI algorithms that are frequently employed in both detection and response, resulting in risk that is difficult to identify, assess, and ultimately escalate to the program level. The reason for this is that while AI techniques are good at detecting anomalies and (potential) attacks, it is often unclear why such signals are given and how the assigned risk may be interpreted from an algorithmic and risk governance perspective. Furthermore, human in the loop approaches may be necessary, in which case a particular challenge is how to effectively combine AI-enabled systems and human risk assessment and escalation (Buczak & Guven, 2015; Berman et al., 2019). This is particularly the case for SoSs that include a number of different stakeholders.

In addition, because the orchestration or coordination of AI-enabled systems may often require collaboration across (and beyond) organizations or teams with different priorities, risk appetites, and responsibilities, there may be a lack of consensus over how to effectively perform such a function. This may be true in particular of systems-of-systems, which often by design are characterized by a degree of managerial independence and autonomy, with the system-of-interest providing governance with limited delegation of authority and responsibility (Dahmann & Baldwin, 2008). In fact, it may be the case that security-related risk management standards are

not always applied with fidelity in practice (Stoneburner et al., 2002; Blank & Gallagher, 2012), which could pose additional challenges to coordinating cybersecurity and AI-related risk in practice.

7.3 Relevance for Large-Scale and Critical Cyber Infrastructures

Systems-of-systems–based orchestration of risk in AI-enabled systems, as described above, is of particular relevance for large-scale and critical cyber infrastructures, including but not limited to national communication networks and information infrastructures, industrial control systems, large-scale distributed sensor networks, and cyber–physical systems. The main reason for this is that critical infrastructure is often used to describe large-scale systems or infrastructures in which component failures may propagate rapidly and lead to extreme consequences not only in terms of technical damage but also economic or even societal loss (Wang et al., 2006; Schneier, 1999). Program-level orchestration of such risk is therefore essential for effective risk management, which would in fact benefit from coordinated and cross-domain oversight in such cases.

In the context of critical infrastructures, SoS–based orchestration of risk is important for ensuring strategic alignment of security operations, whether at the operational or tactical levels. The reason for this is that system- and program-level insights with respect to potential risk not only provide technical input on AI-enabled systems (e.g., in terms of vulnerabilities or anomalies), but also support strategic risk governance by linking security and performance-related considerations with respect to such systems to executive decision-making and a wide range of potential trade-offs, such as between increased resilience, performance, and associated costs (Gordon & Loeb, 2002; Anderson & Moore, 2006). In addition to an immediate effect on resilience, the described orchestration of risk in AI-enabled systems could support sustainability by enabling learning and adaptation in cyber defense programs.

8. Conclusion and Future Research Directions

8.1 Summary of Theoretical and Managerial Contributions

This study advances cybersecurity and risk management scholarship by reframing AI-enabled cyber infrastructure as a system-of-systems that requires coordinated, program-level oversight rather than isolated, component-based controls. Theoretically, the paper integrates three traditionally fragmented bodies of literature: AI-based cybersecurity mechanisms, systems-of-systems engineering principles, and standardized risk management frameworks. By doing so, it extends systems-of-systems theory into the domain of cyber risk governance, building on foundational architectural principles articulated by Maier and later expanded by Jamshidi and Dahmann and Baldwin.

From a risk management perspective, the proposed framework moves beyond conventional risk aggregation models described in established standards and guidelines by introducing the concept

of risk orchestration. This contribution is significant because it recognizes that cyber risks in AI-enabled environments are not merely additive but emergent, adaptive, and often non-linear, as highlighted in prior work on anomaly detection and machine learning limitations in operational settings (Sommer & Paxson, 2010; Garcia-Teodoro et al., 2009).

Managerially, the study provides a structured decision-support perspective that aligns technical cybersecurity signals with governance, accountability, and economic rationality. By explicitly incorporating insights from information security economics, the framework supports more efficient allocation of cybersecurity resources at the program level, addressing the long-standing challenge of diminishing returns in security investments identified by Gordon and Loeb and Anderson and Moore. This alignment enables senior decision-makers to view AI cybersecurity not solely as a technical function but as a strategic, enterprise-level risk management concern.

8.2 Importance of Program-Level Risk Orchestration for Resilient AI Cybersecurity

The findings of this study underscore the critical importance of program-level risk orchestration in achieving resilience within AI-enabled cyber infrastructures. As AI-driven intrusion detection and response systems become increasingly interconnected through shared data, models, and operational platforms, localized risk controls are insufficient to prevent cascading failures. Prior research on intrusion detection in wireless sensor networks, IoT, and distributed systems has already demonstrated that vulnerabilities in one subsystem can rapidly propagate across networked environments (Wang et al., 2006; Zarpelão et al., 2017; Butun et al., 2013).

Program-level risk orchestration addresses this challenge by enabling coordinated visibility, prioritization, and response across autonomous cybersecurity subsystems. Instead of optimizing individual detection engines in isolation, orchestration emphasizes collective resilience, ensuring that risk signals identified in one part of the system inform mitigation strategies across the entire program. This approach is particularly important in AI-driven environments, where learning models can exhibit opaque behavior and performance degradation over time, potentially masking emerging threats until they escalate into systemic failures (Ferrag et al., 2020; Berman et al., 2019).

Moreover, by aligning orchestration with standardized risk governance principles, the framework supports consistency, accountability, and auditability across complex cyber programs. This is especially relevant for large-scale and mission-critical infrastructures, where fragmented risk ownership can undermine both security effectiveness and organizational trust. Program-level orchestration therefore represents a necessary evolution in cybersecurity management, bridging the gap between advanced AI capabilities and robust, system-wide risk governance.

8.3 Need for Empirical Validation and Longitudinal Case Studies

While this study provides a strong conceptual foundation, it also highlights the need for future empirical research to validate and refine the proposed program-level risk orchestration framework. Existing cybersecurity research has largely relied on controlled experiments, benchmark datasets, and simulation-based evaluations of AI models, which, although valuable, offer limited insight into long-term risk dynamics in operational environments (Apruzzese et al., 2018; Buczak & Guven, 2015).

Future research should therefore focus on longitudinal case studies of AI-enabled cyber infrastructures operating across extended time horizons. Such studies would allow researchers to observe how risks evolve as systems adapt, models drift, and organizational priorities change. Empirical investigations could also assess how effectively orchestration mechanisms mitigate cascading failures compared to traditional, siloed risk management approaches.

In addition, cross-sector comparative studies would be valuable in examining how program-level risk orchestration performs in different operational contexts, such as enterprise IT environments, critical infrastructure systems, and large-scale sensor or IoT networks. Integrating qualitative governance analysis with quantitative risk and performance metrics would further strengthen the evidence base for orchestration as a practical management paradigm.

Overall, advancing empirical validation efforts will be essential for transforming program-level risk orchestration from a conceptual framework into a rigorously tested and operationally grounded approach to managing AI-enabled cyber risk.

References

1. Wang, Y., Attebury, G., & Ramamurthy, B. (2006). A survey of security issues in wireless sensor networks.
2. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018, May). On the effectiveness of machine and deep learning for cyber security. In 2018 10th international conference on cyber Conflict (CyCon) (pp. 371-390). IEEE.
3. Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. Information, 10(4), 122.
4. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419.
5. Zarpelão, B. B., Miani, R. S., Kawakani, C. T., & De Alvarenga, S. C. (2017). A survey of intrusion detection in Internet of Things. Journal of Network and Computer Applications, 84, 25-37.
6. Butun, I., Morgera, S. D., & Sankar, R. (2013). A survey of intrusion detection systems in wireless sensor networks. IEEE communications surveys & tutorials, 16(1), 266-282.

7. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. computers & security, 28(1-2), 18-28.
8. Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications surveys & tutorials, 18(2), 1153-1176.
9. Sommer, R., & Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE symposium on security and privacy (pp. 305-316). IEEE.
10. Klein, J., & Van Vliet, H. (2013, June). A systematic review of system-of-systems architecture research. In Proceedings of the 9th international ACM Sigsoft conference on Quality of software architectures (pp. 13-22).
11. Jamshidi, M. (Ed.). (2017). Systems of systems engineering: principles and applications. CRC press.
12. Dahmann, J. S., & Baldwin, K. J. (2008, April). Understanding the current state of US defense systems of systems and the implications for systems engineering. In 2008 2nd Annual IEEE Systems Conference (pp. 1-7). IEEE.
13. Schneier, B. (1999). Modeling security threats. Dr. Dobb's journal, 24(12).
14. Anderson, R., & Moore, T. (2006). The economics of information security. science, 314(5799), 610-613.
15. Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. ACM Transactions on Information and System Security (TISSEC), 5(4), 438-457.
16. Lalonde, C., & Boiral, O. (2012). Managing risks through ISO 31000: A critical analysis. Risk management, 14(4), 272-300.
17. Purdy, G. (2010). ISO 31000: 2009—setting a new standard for risk management. Risk Analysis: An International Journal, 30(6), 881-886.
18. Blank, R. M., & Gallagher, P. D. (2012). Guide for conducting risk assessments. NIST Special Publication, 800(30).
19. Stoneburner, G., Goguen, A., & Feringa, A. (2002). Risk management guide for information technology systems. Nist special publication, 800(30), 800-30.
20. Maier, M. W. (1998). Architecting principles for systems-of-systems. Systems Engineering: The Journal of the International Council on Systems Engineering, 1(4), 267-284.