Federal Register Notice: 89 FR 78915, Federal Register: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Public Input on the National Cyber-Physical Systems Resilience Plan

On behalf of the Office of Science and Technology Policy (OSTP), the NITRD National Coordination Office issued cBrequest for information (RFI) (89 FR 78915) seeking public input for the creation of a National Plan for Cyber-Physical Systems Resilience Research. This RFI was issued in response to the February 2024 report by the President's Council of Advisors on Science and Technology (PCAST), which recommended that NITRD "formulate a more coordinated national research and development (R&D) agenda on cyber-physical resilience".

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Table of Contents

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Alina Zare, Ph.D.

Zoleikha Biron, Ph.D.

Swarup Bhunia, Ph.D.

Sara Rampazzi, Ph.D.

Sandip Ray, Ph.D.

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

UF UNIVERSITY *of* FLORIDA

Herbert Wertheim College of Engineering
Associate Dean of Research and Facilities

[redacted]

Subject: Response to NSF RFI 2024-22005 - 'Cyber-Physical Systems Resilience R&D Plan'

Dear Suzanne H. Plimpton,

I am writing on behalf of the University of Florida, a leader in cyber-physical systems, resilience, and cybersecurity research, to submit our response to your recent Request for Information (RFI), NSF RFI 2024-22005 - 'Cyber-Physical Systems Resilience R&D Plan'. We are excited to have this opportunity to engage with the NSF community and share our expertise in addressing the pressing need for building Cyber-Physical System Resilience.

Our attached response to your RFI poses insightful suggestions, innovative ideas, and perspectives on resilience of critical infrastructure in the face of climate change, natural disasters or harsh environments (such as operating in the deep sea or in space). Our team has deep expertise in the relevant fields, with extensive experience in the research and development of cyber-physical systems and security of these systems. We also have research groups studying the resilience of our natural systems, coastal environments, and physical infrastructure in the face of climate change and several natural disasters (with a particular emphasis on hurricanes). We thoroughly understand the complexity of fortifying these systems while taking into account technical as well as social considerations.
In conclusion, we are excited to discuss our response further with you and look forward to future opportunities for research collaboration.

Sincerely,

[redacted]

Alina Zare, Ph.D.
Associate Dean of Research and Facilities, Herbert Wertheim College of Engineering
Professor, Electrical and Computer Engineering
Director, Machine Learning and Sensing Lab
University *of* Florida
a[redacted]

University of Florida’s Response to NSF RFI 2024-22005 - 'Cyber-Physical Systems Resilience R&D Plan'

1. **Resiliency Challenges and Approaches in Critical Infrastructure Under Adverse Environmental Conditions due to Natural Disasters and Climate Change**
Climate change is a stark and impending reality of our global environment. The very survival of critical cyber-physical system (CPS) infrastructure depends on our ability to be resilient against environmental impacts induced by climate change. It should be a research priority to determine (1) the impact of climate change on the functionality, safety, and sustainability of critical infrastructure and (2) strategies to make them resilient. Specifically, the transportation and energy sectors can be highly susceptible to the impacts of climate change in two ways. First, these infrastructures must achieve long-term sustainability in the transformed environment, e.g., a roadway or bridge must be resilient to higher temperature swings. Research in these areas marries the exploration of new materials (e.g., materials more resistant to extreme conditions) and research in AI and sensor technology (e.g., to identify aging or weakening infrastructure to enable effective and safe mitigation. Additionally, this research will heavily depend on exploring different resiliency strategies before deploying them in the infrastructure. One way to achieve this is through the practical design of digital twins for critical infrastructure that enables comprehension of their interaction with newer climate models.

    Many researchers at the University of Florida are at the frontier of observing and responding to natural disruptions, e.g., hurricane impact on built-, natural- and cyber-infrastructures. Much of this research area aims to encourage potential paths for novel and applicable technologies that enhance the resiliency of cyber-physical systems (CPSs) and significantly impact human life, society’s revenue, well-being, and social welfare. The resiliency of CPSs toward natural disasters like hurricanes, wildfires, and earthquakes involves multiple strategies and technologies designed to withstand, adapt to, and recover from such events. Most existing research directions focus on preventing and minimizing the impacts of disasters. However, the stage of recovery and adaptable recovery of CPSs against abnormalities is as essential as restoring regular operation for CPSs. Such recovery strategies require localization of damages in widespread cyber-infrastructure, such as power systems, that may take days to finalize and stabilize.

    AI-based CPS resiliency during natural disasters, such as hurricanes needs to leverage advanced technologies to enhance preparedness, response, and recovery. However, most AI-based techniques require large training datasets. Yet, AI-based recovery and resiliency toward natural disasters suffer from limited datasets or facilities equipped with high-fidelity simulators that can simulate the event properly. Investigations to advance these techniques to develop a comprehensive research center that can simulate natural disasters for CPSs with exceptionally reliable models and develop adaptable AI-based resiliency approaches that learn from experiences are necessary. Such technology will not only enhance the resiliency of CPS but will also go beyond existing resiliency methods and make the recovered system wiser and more intelligent toward similar scenarios. Enhancing the resiliency of cyber-physical systems against hurricanes requires a multi-faceted approach that includes robust design, real-time monitoring, predictive analytics, and effective emergency planning. By integrating these elements, organizations can improve their capacity to withstand natural disasters and recover more swiftly, minimizing disruption to critical services and infrastructure.

2. **Resiliency of CPS Systems in New Frontiers including, Deep Sea and Space**

The recently issued PCAST report, *Strategy for Cyber-Physical Resilience: Fortifying our Critical Infrastructure for a Digital World,* appropriately focuses on long-term improvements in critical infrastructure security, focusing on securing current national infrastructures such as Pipelines, Rail, Aviation, Water Systems, Ports, and Hospitals. All these areas merit research investment as these CPSs were not originally built with today's sophisticated AI-driven cyber-threats in mind, making them vulnerable to attacks that take advantage of intrinsic design vulnerabilities, legacy technology, and flaws in software and hardware domains beyond traditional network attacks.

AI-driven threats demonstrate adaptability even in the presence of incomplete or missing data, while defense systems are incredibly challenging to deploy and update on existing old infrastructures. We believe research priority should be given to attack-agnostic detection-reaction technologies and sematic-based approaches to analyze the behavior and relationships of data, AI models input-output relationships, and process flows to identify inconsistencies and determine the root cause of wrong automated and autonomous decisions.

The report does not, however, consider the security of emerging critical infrastructures deployed in isolated environments, such as deep sea and space. New underwater computing and storage platforms, commercial and military spacecraft, satellite swarms, and stations are a growing part of the global critical infrastructure, which deserves significant investigation.

Unlike traditional in-land infrastructures, such sparsely distributed, autonomous networks of moving and stationary systems are exposed to new threats that leverage the uniqueness of the environments in which they are deployed. University of Florida researchers have demonstrated that conventional security practices are insufficient to prevent data alteration or disruption of such CPS. We believe research is needed to revisit zero-trust architectures, enhance hardware and software side-channel resiliency, improve sensing and actuation technology robustness against novel physical and digital spoofing, advance autonomous security, ensure secure data storage and recovery, and maintain data integrity control over heterogeneous communication channels and across different physical mediums.

Research is also needed on innovations in tools and frameworks for rigorously testing software, hardware, and data security requirements during the design phase of such novel CPSs, which must withstand harsh environments and sophisticated attackers for prolonged periods without the possibility of immediate human intervention.

Summary of Response

Climate change poses significant risks to critical cyber-physical systems (CPS) infrastructure, especially in the transportation and energy sectors, necessitating research to enhance their resilience. The focus should be on understanding how climate change affects infrastructure functionality and developing innovative materials and AI technologies to improve safety and sustainability.

At the University of Florida, researchers are exploring strategies for CPS resilience against the increasing prevalence of natural disasters like hurricanes, emphasizing the importance of recovery processes alongside prevention. AI technologies are envisioned to enhance preparedness, response, and recovery, but require investment into systems that can generate simulated training data for training these AI technologies.

The PCAST report highlights the need for improved security in existing infrastructures, which were not designed for modern AI-driven cyber threats. It stresses the importance of developing attack-agnostic technologies to detect and analyze vulnerabilities. However, it overlooks the security of emerging critical infrastructures in isolated environments, such as deep sea and space, which are becoming increasingly significant. A comprehensive approach combining robust design, real-time monitoring, predictive analytics, and strategic emergency planning is essential to enhance CPS resilience against both natural disasters and cyber threats.

Key Findings and Recommendations

Key Findings:

1. **Climate Change Impact:** Critical infrastructure, particularly in transportation and energy, is highly vulnerable to the effects of climate change, necessitating research on functional sustainability.
2. **Resilience Strategies:** Existing research often focuses on disaster prevention, but effective recovery and adaptive strategies are equally important.
3. **AI Limitations:** Current AI-based approaches for enhancing CPS resilience face challenges due to limited training datasets and the need for high-fidelity simulators for disaster simulation.
4. **Cyber Threat Vulnerability:** Many existing CPS infrastructures were not designed to withstand modern AI-driven cyber threats, making them susceptible to attacks exploiting inherent vulnerabilities.
5. **Emerging Infrastructures:** Security considerations for new CPS in isolated environments like deep sea and space are currently underexplored.

Recommendations:

1. **Research Prioritization:** Invest in research to assess the impact of climate change on critical infrastructure and identify resilience strategies.
2. **Enhancing Recovery Strategies:** Develop comprehensive recovery plans that focus on rapid damage localization and adaptable recovery processes in CPS.
3. **Advanced AI Techniques**: Create improved AI models that can learn from past experiences and effectively simulate natural disasters, ensuring they are equipped with extensive training datasets.
4. **Attack-Agnostic Solutions:** Focus on developing detection-reaction technologies that analyze data relationships and identify inconsistencies, enhancing security against evolving cyber threats.
5. **Explore Isolated Environments:** Increase research attention on the security of critical infrastructures in isolated environments, including underwater and space systems, acknowledging their growing significance.

By addressing these points, organizations can enhance the resilience and security of critical CPS against both environmental and cyber threats.

Conclusion
We hope that our response has provided valuable information and insights that can inform future research directions and decisions at the NSF. We look forward to the opportunity to discuss our response further and explore potential research collaborations.

Sincerely,
Zoleikha Biron, Ph.D.
Swarup Bhunia, Ph.D.
Sara Rampazzi, Ph.D.
Sandip Ray, Ph.D.
Alina Zare, Ph.D.
University of Florida

Notes: Dr. Kevin Butler recused himself from this response because he is on the PCAST Review Working Group. This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Alvaro A. Cardenas

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Cyber-Physical Resilience through Autonomous Attack-Recovery.

Alvaro A. Cardenas
UC Santa Cruz

Security is a process that includes prevention, detection, and response to attacks. Incident response, the last of these steps, is significant when considering cyber-physical systems due to their real-time constraints and safety risks.

An incident response strategy usually requires a plan for collecting and keeping logs of sensitive events, analyzing them to identify the causes, and giving them a priority level. Response strategies then focus on containing the attack, closing vulnerabilities, eradicating the threat, and recovering the system (e.g., re-imaging the affected system). These actions usually involve human analysts and can take hours or even days until the system is fully restored.

While all these actions are needed in an offline review of an attack, in cyber-physical systems we have real-time requirements that cannot wait until an offline review of the event is completed. For example, a physical process under attack such as a chemical reactor with increasing dangerous internal pressure may cause an explosion before a security analyst can detect and understand what is going on. Similarly, an autonomous vehicle under attack may crash or run over pedestrians before a remote operator can step in. Therefore, in addition to an offline incident response plan, we also need a plan to mitigate attacks promptly by developing online attack recovery algorithms that keep the system safe and allow it to complete its mission.

Our recent work on attack-recovery [1] tries to address this gap. However, many challenges remain. While our previous efforts can help prevent immediate safety risks, they still require the designers of cyber-physical systems to identify several parameters before the operation of the system, such as safe states to drive the system to (targets) after an attack is detected; and thus they are not adaptable to uncertain conditions and new attacks.

To address these limitations we need more research leveraging new advances in autonomous common-sense reasoning, to help automate the recovery of cyber-physical systems [2]. In particular, the state-of-the-art automatic attack-recovery mechanisms do not work with dynamic and uncertain environments. For example, these previous methods need precomputed target safe areas where the recovery controller can take the system; however, if these sets are not preloaded in advance, or if the safe zones are not ``safe'' at the time, the automatic recovery mechanism will fail.

To address the limitations of previous attack recovery systems, we need more research on AI to develop an autonomous-based attack recovery mechanism. Our main insight is to have a hierarchical recovery strategy --- At the lower level we will use mathematical control-theory models based on the simplex architecture, and at a higher level, we need to design an AI recovery algorithm to provide a common-sense and adaptive recovery plan.

As exemplified in [2], after detecting an attack against a drone flying over a city, we can ask an AI agent to identify a safe area where the drone can land (given the camera feed of the drone). The agent must decide which building the drone should land on. The agents need to assess the risk of an emergency landing in all possible locations and then predict several candidate safe landing zones based on both the original and preprocessed views. Each predicted landing zone includes coordinates and a ``Reason'' section to improve prediction accuracy and interoperability.

In short, future autonomous systems need to have fail-safe conditions that are adaptive to dynamical and unpredicted conditions. We need further research for autonomous attack recovery that is efficient and secure.

References

[1] Zhang L, Burbano L, Chen X, Cardenas AA, Drager S, Anderson M, Kong F. Fast Attack Recovery for Stochastic Cyber-Physical Systems. In 2024 IEEE 30th Real-Time and Embedded Technology and Applications Symposium (RTAS) 2024 May 13 (pp. 280-293). IEEE.

[2] Ortiz D, Burbano L, Yang S, Wang Z Cardenas AA, Xie C, Cao Y. Robust and Efficient AI-Based Attack Recovery in Autonomous Drones. In Proceedings of the GENZERO workshop. 2024 November.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Resilience by Design Fault Model Design Tools

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Resilience-By-Design with fmdtools

Resilience, the ability of a system to dynamically mitigate hazardous events as they arise, is an important consideration for a wide range of complex engineered systems across domains. While engineering resilience has traditionally been seen as property of infrastructure (e.g., [1]), the resilience perspective is increasingly having relevance to a wide range of technologies. In aviation, for example, resilience is considered an aspect of human pilots to proactively prevent and perform contingency management actions to recover from hazardous flight events [2], which is becoming increasingly important to replicate in the implementation of autonomous and partially autonomous flight systems [3]. Cases like this especially increase the importance of incorporating resilience in the design of the system prior to implementation, so that new technologies will fulfill the public's high expectations of safety and dependability and not pose undue risks on society.

The fmdtools (Fault Model Design Tools) library[1] was developed to enable the consideration of resilience in the design process [4]. To enable this, it provides modelling constructs to represent a system's structure and (both nominal and hazardous) behavior, methods to simulate the dynamic effects of hazardous scenarios over time, and visual and statistical methods to analyze and thus improve resilience. As a python-based library, fmdtools is widely adaptable across domains to a wide range of use-cases including power systems, drones, and autonomous rovers, as well as system-of-systems and infrastructure use-cases like aerial firefighting [5] and airport taxiways [6].

The development of fmdtools features and capabilities for resilience analysis have been well-documented in the literature, including:

- Representation of human-oriented, automation-oriented, and general systems of systems resilience properties, including task execution [7], performance degradation [8], and distributed situation awareness [6].
- Support for trade assessment and optimization via model parameterization [9]
- Ability to construct and sample large spaces of hazardous scenarios [10]

In the future, it is the goal of the fmdtools developers to make these capabilities more accessible to a wider range of users, through the development of an integrated graphical user interface.

Bibliography

[1] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace and D. Von Winterfeldt, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthquake Spectra,* pp. 733-752, 2003.

[1] Available at: https://github.com/nasa/fmdtools

[2] J. B. Holbrook, M. J. Stewart, B. E. Smith, L. J. Pinzel, B. L. Matthews, I. Avrekh, C. T. Cardoza, O. C. Ammann, V. Adduru and C. H. Null, "Human performance contributions to safety in commercial aviation," NASA, Hampton, Virginia, 2019.

[3] National Academies of Sciences, Engineering, and Medicine, "Advancing Aerial Mobility: A National Blueprint," The National Academies Press, Washington, DC, 2020.

[4] D. Hulse, H. Walsh, A. Dong, C. Hoyle, I. Tumer, C. Kulkarni and K. Goebel, "fmdtools: A fault propagation toolkit for resilience assessment in early design," *International Journal of Prognostics and Health Management,* 2021.

[5] S. R. Adrade and D. E. Hulse, "Evaluation and Improvement of System-of-Systems Resilience in a Simulation of Wildfire Emergency Response," *IEEE Systems,* vol. 17, no. 2, pp. 1877 - 1888, 2022.

[6] L. Irshad and D. Hulse, "Modeling Distributed Situation Awareness in Resilience-Based Design of Complex Engineered Systems," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,* vol. 87295, p. V002T02A050, August 2023.

[7] L. Irshad, D. Hulse, H. O. Demirel, I. Y. Tumer and D. C. Jensen, "Quantifying the Combined Effects of Human Errors and Component Failures," *ASME Journal of Mechanical Design,* p. 101703, 2021.

[8] D. Hulse and L. Irshad, "Using Degradation Modeling to Identify Fragile Operational Conditions in Human- and Component-driven Resilience Assessment," in *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*, Portsmouth, VA, 2022.

[9] D. Hulse and C. Hoyle, "Understanding Resilience Optimization Architectures: Alignment and Coupling in Multilevel Decomposition Strategies," *ASME Journal of Mechanical Design,* p. 111704, 2022.

[10] D. Hulse and L. Irshad, "Synthetic Fault Mode Generation for Resilience Analysis and Failure Mechanism Discovery," *ASME Journal of Mechanical Design,* p. 031707, 2023.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Anonymous

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

10/25/2024; anonymous and received via FDMS

Biometrics don't work. Refer to the latest USCCR report on this, about facial recognition. The official recommendation is to NOT use biometrics because they have very high error rates among POC and the elderly, and there are currently no biometric data use policies posted on the DHS website to deter privacy concerns. See https://www.usccr.gov/reports/2024/civil-rights-implications-federal-use-facial-recognition-technology With respect to FRT accuracy and bias, the National Institute of Standards and Technology (NIST) testing is voluntary and represents laboratory—not real-world—results. Thus, NIST cannot say that its evaluated programs are accurately representative of the performance of all FRT deployed throughout the country. Algorithmic accuracy rates can vary widely among developers, but even with the highest-performing algorithms, tests have shown there are likely to be false positives for certain demographic groups, specifically Black people (particularly Black women), people of East Asian descent, women, and older adults. A promising FRT testing model does exist: DHS, through its Science and Technology Directorate, funds FRT research, testing, and evaluation at MdTF, which specializes in “scenario testing” of the entire FRT system as it is intended to be deployed. DHS is the only agency known to be testing FRT in this way. Any agency using FRT should have a publicly available use policy. If agencies do use FRT, they should audit their use to ensure it complies with government policy. FRT vendors providing the federal government with solutions should provide users with ongoing training, technical support, and software updates to ensure their systems can maintain high accuracy across demographic groups in real-world deployment contexts. Furthermore, agencies should ensure their CAIOs work in close coordination with existing responsible officials and organizations within their organizations, including Civil Rights and General Counsel offices, to advise and update agency FRT guidance, implementation, and oversight. Federal grantees using FRT should provide verified results with respect to accuracy and performance across demographics from NIST’s FRT Evaluation or similar government-validated third-party test.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Bethany Johns

Jacob Dowd

Monowar Hasan

Assefaw Gebremedhin

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

WASHINGTON STATE UNIVERSITY

Office of External Affairs and Government Relations

Protecting Digital Agriculture from Cyber Breaches

RFI Response: *Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience*

Bethany Johns
Assistant Director, Federal Relations, Washington State University
[redacted]

Jacob Dowd
Director, Federal Relations, Washington State University
Email: [redacted]

Monowar Hasan
Assistant Professor, School of EECS, Washington State University
Email: [redacted]

Assefaw Gebremedhin
Associate Professor, School of EECS, Washington State University
Email: [redacted]

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

RFI Response Focus: The U.S. agriculture sector is being infused with various cyber components. However, the cyber-physical resilience of these digital assets from threats ranging from extreme weather, power grid outages, to cyberattacks is still an afterthought. There is a need for a critical analysis of the cybersecurity posture of precision agriculture applications and improving the preparation and protection of this sector against cyberattacks. Protecting the U.S. agriculture sector is critical and requires a concerted, cross-domain research, thoughtful synthesis of ideas and methods, and investment in workforce development.

1. Introduction

Increased reliance on digital technologies and networked distributed systems across various industrial sectors has made critical infrastructure cybersecurity one of the top priorities for national security. The U.S. food and agriculture sector continues to be identified as one of the 16 critical national infrastructures in the U.S., accounting for roughly one-fifth of the nation's economic activity [1], [2]. As this sector adopts advanced and innovative digital technologies -- such as connected sensors, embedded computers, smart tractors, drones and Artificial Intelligence (AI) - productivity has increased, nutritional value has improved, disease resistance, and resilience to climate change. While these technologies have enabled unprecedented gains in innovation and productivity, they expose modern agriculture to various cyber vulnerabilities impacting precision agriculture operations. Due to the high value of agricultural sectors to adversaries, limited or no cybersecurity defense, and lack of domain-specific security understanding of those systems, digital agricultural systems are increasingly becoming targets for cyber breaches. The food and agricultural sectors have faced a rising number of cyberattacks, with over 30 major incidents have been reported [6], Recent examples of major cyber incidents in the U.S. agricultural sector include ransomware attacks on JBS facilities [3], Russian hackers targeting an Iowa grain co-op [4], and a Minnesota grain handler [5]. leading to significant financial losses. Attacks originating from other critical sectors, such as the utility grid and industrial control systems, could also threaten the food supply chain due to close interdependence, as indicated by the U.S. Department of Energy [7]. Any cyberattack targeting agriculture and closely related critical infrastructure could jeopardize the nation's agricultural production, exports, food security, and ultimately, national security.

A critical limitation in our progress on security for cyber-physical systems is the lack of knowledge and training of professionals in the domains these systems are being deployed. For example, there is a growing need for individuals with knowledge in the agriculture, food, and life sciences domains to help build resilient cyber systems that can withstand threats to these domains. There is much within the food and agriculture knowledge domain that requires specialized threat intelligence to mitigate threats and ensure a resilient system. Cross-domain knowledge will become increasingly vital as cyber-physical systems intelligently interact with other systems across information and physical interfaces.

2. Research Needs

The following systematic research is needed for a thorough understanding on how resilience by design or resilience by intervention can prepare for recovery and adaptation in different threat scenarios.

Risk Assessment and Cyber Forensics Tool. Development of new software tools and cyber risk models for optimal cybersecurity investment, enabling ag asset owners to assess risks and implement cost-effective mitigations continuously. Resilience strategies will include redundant systems for critical operations, automated recovery protocols, and fallback mechanisms to maintain safety, such as livestock health, even during system compromises. A comprehensive set of "what-if" contingency scenarios need to be developed with optimal precomputed remedial actions to address potential operational impacts, safety concerns, and economic losses.

AI and Machine Learning-based Anomaly Detection and Mitigation. Incorporating AI and machine learning, deep learning, and reinforcement learning for automated detection and real-time mitigation of cyber anomalies. Federated learning will enable collaboration among distributed agents for system-wide intelligence while preserving privacy. How can such models continuously learn and adapt to evolving threats and optimize detection strategies based on domain-specific data? How do we integrate those models into Security Information and Event Management (SIEM) platforms for real-time cybersecurity solutions?

Zero-Trust Platform. To support attack prevention, mitigation, and cyber resilience, how do zero-trust principles (i.e., "never trust, always verify") and moving target defense algorithms adapt system parameters to reduce attack surface and prevent attacks dynamically? The food and ag sectors have a combination of Information Technology (IT) and Operation Technology (OT) systems, and the forensic analysis in OT presents additional challenges as they often do not capture system state (via system logs or network logs) coupled with legacy hardware and embedded software still being utilized in these systems. We need to develop cyber forensics tools for food and agriculture OT systems, focusing on recording state information and conducting deep log analysis. One challenge is identifying threat intelligence (attack types, sources, compromised subsystems, or credentials) so that a zero-trust model can be effective.

Blockchain-based Data Sharing Platform. Can distributed ledger technology support product traceability and secure data sharing, with robust algorithms ensuring data security and privacy? Malicious alterations to sensitive farm data can cause financial losses and harm to crops or livestock. The challenge is to study whether traditional privacy techniques are sufficient for the evolving agricultural ecosystem.

Systems Integration. Further research is needed on integrating developed cybersecurity technology testbeds into real-world cyber secure and resilient system testbeds for the food and ag sector, namely, the animal agriculture testbed, food crops testbed, agricultural machinery testbed, and inter-dependent agriculture-water-energy testbed. Synergistic partnerships with various ag stakeholders and industry, government, national labs, and academic institutions are needed to build and release real-world datasets and use-case scenarios. The developed testbeds can then use the datasets and attack vectors to demonstrate their feasibility and efficacy concerning cybersecurity.

Agriculture Cybersecurity and Climate Change. Cybersecurity in agriculture directly impacts how effectively the sector can mitigate and adapt to the effects of climate change. Modern agricultural practices depend on data-driven technologies such as satellite-based weather forecasting, precision irrigation, and climate modeling. These technologies enable farmers to make informed decisions about crop management, water usage, and soil health, which are crucial for reducing agriculture's carbon footprint and conserving water in the face of climate variability. If cybersecurity vulnerabilities in these systems go unchecked, they could be exploited to disrupt data flows or corrupt decision-making models. For example, an attack that manipulates climate or weather data could lead to incorrect decisions about irrigation scheduling, pesticide application, or crop planting, ultimately undermining efforts to adapt to changing environmental conditions. Another challenge is isolating unintentional faults from cyberattacks so farmers/ag-tech vendors can proactively make informed decisions. Hence, there is a need for a new data-driven model and leverage advancement of AI technologies to understand the impact of climate change on agriculture decision-making. Therefore, protecting agricultural systems from cyber threats is integral to ensuring that the sector can continue to play a vital role in addressing climate change through sustainable and efficient practices.

3. Cross-Domain Exploration

Agriculture does not exist in isolation. It is deeply intertwined with other critical infrastructure sectors, such as the power grid, water management, and transportation systems. As a result, cybersecurity vulnerabilities in other domains can have cascading effects on agriculture, jeopardizing food security and economic stability. For instance, a cyberattack on the power grid or energy delivery system can directly impact agriculture by disrupting power to irrigation systems, sensors, and storage facilities. This can lead to crop loss, water waste, and reduced resiliency in agricultural operations, especially during extreme weather events. A failure in the power grid could also affect the cold chain logistics critical for preserving food post-harvest, leading to significant food spoilage and supply chain disruptions. For instance, if hackers were to compromise the data networks used in energy distribution, they could manipulate electricity prices or availability in ways that severely impact the cost and reliability of energy needed for farming equipment, heating systems for livestock, or greenhouses.

Similarly, cybersecurity vulnerabilities in transportation systems could delay the delivery of essential supplies such as seeds, fertilizers, and machinery or prevent crops from reaching markets on time, affecting food availability and prices. **We stress that cybersecurity research in agriculture must consider how interrelated sectors like the power grid, water infrastructure, and logistics are interconnected.** This cross-domain research is essential for developing comprehensive security strategies that ensure agriculture can remain resilient despite cascading disruptions from other sectors.

4. Workforce Development and Interdisciplinary Research Environment

There is a critical need for a well-trained and diverse cybersecurity workforce in the US, across all sectors. Federal initiatives to cultivate the US cyber workforce, such as the Workforce Framework for Cybersecurity, National Centers of Academic Excellence in Cybersecurity, and the 8140 series of Department of Defense (DoD) Directives, have yielded much dividend. A recent addition to these initiatives is the establishment of the Virtual Institutes for Cyber and Electromagnetic Spectrum Research and Employ (VICEROY). The mission of VICEROY is to "Establish academic cyber institutes at institutions of higher learning, to develop foundational expertise in critical cyber and spectrum operational skills for future military and civilian leaders of the Armed Forces and the DoD, including such leaders of the reserve components." Towards this goal, VICEROY funds undergraduate cybersecurity education at designated Virtual Institutes.

One such institute is the VICEROY Northwest Institute for Cybersecurity Education and Research (CySER) [8]. Established in 2021 in the first cycle of VICEROY, CySER is a consortium of four institutions in the Northwest led by Washington State University (WSU) with partner institutions Montana State University (MSU), University of Idaho (UI), and Central Washington University (CWU). CySER's focus is training a Reserve Officer Training Corps (ROTC) and national defense-oriented civilian workforce, combining cybersecurity technical competency with professional competencies of leadership, communication, teamwork, and lifelong learning. The CySER program has thus far had remarkable achievements, including catalyzing the launching of a new Bachelor of Science in Cybersecurity degree program at WSU and increasing collaboration in research and education across multiple institutions [9].

Federal investment in programs like VICEROY is needed in sectors outside of defense as well, including in agriculture. Institutions such as WSU, where a successful VICEROY program has existed and is well established, serve as valuable models for scaling up and branching out cybersecurity workforce training into other domains.

A matter closely tied to the education and training of cybersecurity workforce is the **establishment of reliable cyber ranges** that students and other trainees can easily access. Cyber ranges are interactive and simulated platforms that replicate networks, systems, tools, and applications. They provide a safe and legal environment for acquiring hands-on cyber skills, and they have a crucial role to play in cybersecurity education and training. They may consist of actual hardware and software or a combination of physical and virtual components. Types of cyber ranges commonly used include *simulation ranges*, *overlay ranges*, *emulation ranges*, and *hybrid ranges*. The type that would be most desirable for contexts in academic institutions such as WSU is a hybrid range, which would integrate features and capabilities from simulations, overlay ranges, and emulation, creating a hybrid environment that suits specific requirements.

5. Beyond Agriculture

Considering the diverse agriculture domain (plants, crops, livestock, seafood) and external climatic variabilities (weather, water), understanding cybersecurity issues and building techniques using digital twins to bolster security posture requires concerted, multiyear efforts.

Although the majority of this RFI response document focused on agricultural cybersecurity, accompanying research is needed to investigate how this agricultural domain knowledge can be transferred to understanding and protecting other critical sectors, such as power grids, transportation systems, biomanufacturing, public health, food supply chain, and wildfire management, using the digital twin technologies.

References

[1] "GIAC Cyber Security Discussion Paper." Accessed: Jun. 06, 2024. [Online]. Available: https://www.ams.usda.gov/about-ams/giac-may-2024-meeting/cybersecurity

[2] "Food and Agriculture Sector-Specific Plan - 2015." Accessed: Jun. 06, 2024. [Online]. Available: https://policycommons.net/artifacts/12474227/food-and-agriculture-sector-specific-plan/13370861/

[3] "JBS S.A. ransomware attack," Wikipedia. Accessed: Jun. 15, 2024. [Online]. Available: https://en.wikipedia.org/wiki/JBS_S.A._ransomware_attack

[4] "BlackMatter Hits Grain Cooperative With Ransomware Attack." Accessed: Jun. 15, 2024. [Online]. Available: https://www.itprotoday.com/attacks-breaches/blackmatter-hits-grain-cooperative-with-ransomware-attack

[5] "Minnesota grain handler targeted in ransomware attack." Accessed: Jun. 15, 2024. [Online]. Available: https://www.reuters.com/technology/minnesota-grain-handler-targeted-ransomware-attack-2021-09-23/

[6] A. Kulkarni, Y. Wang, M. Gopinath, D. Sobien, A. Rahman, and F. A. Batarseh, "A Review of Cybersecurity Incidents in the Food and Agriculture Sector," Mar. 12, 2024, *arXiv*: arXiv:2403.08036. Accessed: Jun. 06, 2024. [Online]. Available: http://arxiv.org/abs/2403.08036

[7] US Department of Energy, "Cybersecurity considerations for distributed energy resources on the US electric grid." [Online]. Available: https://www.energy.gov/sites/default/files/2022-10/Cybersecurity%20Considerations%20for%20Distributed%20Energy%20Resources%20on%20the%20U.S.%20Electric%20Grid.pdf

[8] "VICEROY Northwest Institute for Cybersecurity Education and Research (CySER)." Accessed: Jun. 05, 2024. [Online]. Available: https://cyser.wsu.edu/

[9] "MCPA - Crabb - Cyber Education and Research — public.milcyber.org." [Online]. Available: https://public.milcyber.org/activities/magazine/articles/2024/crabb-cyber-education-and-research

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Charles Cao

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Comment via FDMS

10/17/2024

Charles Cao, [redacted]

[redacted]

In response to the RFI on Cyber-Physical Systems Resilience, we propose a comprehensive approach to strengthen the resilience of our nation's critical infrastructure through advanced software patching techniques. Our method focuses on rapidly fixing vulnerabilities in existing systems in the binary level firmware by rewriting certain sections, without the need for complete overhauls, which is crucial for maintaining and improving the reliability of older, legacy software that many of our critical systems still rely on. By developing smart, efficient ways to update software in power grids, transportation systems, and industrial controls, we can significantly enhance their ability to withstand and recover from various disruptions, whether caused by cyber attacks, natural disasters, or human error. This approach not only addresses immediate security concerns but also provides a cost-effective way to modernize our infrastructure, ensuring that essential services remain available and reliable for all Americans, even in the face of unexpected challenges.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Computing Community Consortium (CCC)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

CCC
Computing Community Consortium
Catalyst

CCC’s response to the Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience

This response is prepared by the Computing Research Association (CRA)’s Computing Community Consortium (CCC). CRA is an association of over 270 North American computing research organizations, both academic and industrial, and partners from six professional computing societies.

The mission of the CCC, a subcommittee of CRA, is to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges. Please note any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the authors’ affiliations, or of the National Science Foundation, which funds the CCC.

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

October 26, 2024

Written by: *Gabrielle Allen (University of Wyoming), David Danks (University of California, San Diego), Catherine Gill (Computing Community Consortium), and Katie Siek (Indiana University).*

To ensure the National Plan for Cyber-Physical Systems Resilience Research is thorough and advantageous to the nation, the Computing Community Consortium presents the following recommendations.

1. Prioritize Cyber-Physical-Human Resilience Efforts

Resilience of cyber-physical systems is not just about the software or infrastructure on which a system depends; it also involves the people who design, develop, operate, and maintain every aspect of these systems. The national plan should consider human operators as distinct components in cyber-physical resilience planning. Redundancies in cyber-physical systems are key to maintaining a level of operability during crises, and similar redundancies must also be ensured at the human level (e.g., redundancies in knowledge and expertise). Systems also often lack adequate monitoring and sensing on the human side, in contrast with cyber and physical components, even though human decisions and actions can significantly impact these systems' operations. Research should explore how humans interact with systems and the potential for human error, burnout, and adaptation in crisis situations. Research should also be conducted to discover optimal triage methods for a given cyber-physical system during crises to ensure threats are managed quickly, rather than exacerbated, by on-site workers. Monitoring devices and artificial intelligence may be useful to assist humans with time critical decisions, and research into how best to use these technologies should be pursued and included in the plan.

To design cyber-physical-human systems with *all* stakeholders in mind, the strategic plan should emphasize the need to train scholars from historically excluded groups. Based on our experiences, we emphasize that institutions should be required to have checks and balances to ensure people from historically excluded groups are provided with real research experiences and treated ethically. This could be done with comparative pre-, mid-, and post-research experience surveys with one data group to make comparison reports. To further support the pipeline of future researchers, we recommend funding summer research programs for MS students, particularly those in intensive short-term programs, who often lack the opportunity to gain research experience. Moreover, the agency should require funded institutions to document their efforts in fostering an inclusive environment, with clear metrics to assess and improve the institutional culture. By doing so, the agency can shift the responsibility of addressing hostile environments from underrepresented groups to institutional leaders, ensuring that the burden is on institutions to create a supportive and welcoming atmosphere for all cyber-physical systems researchers.

2. Open Source Software and Cyber-Physical Resilience

The strategic plan should carefully evaluate the role of open-source software and tools in cyber-physical systems. Open source software can accelerate innovation, but it also presents unique security challenges. As we've seen recently, even a reputable open-source system used by leading US based companies, significant portions of the US government, and hundreds of millions of computers around the world can be susceptible to cyberattacks[1]. It is important to identify where open-source solutions can be securely integrated into critical infrastructure while ensuring proper testing and

[1] https://www.nytimes.com/2024/04/03/technology/prevent-cyberattack-linux.html

validation of code, including AI-generated software. For open-source systems which are widely used across the US and the world, mechanisms must be introduced to reduce the effectiveness of cyberattacks and staunch the bleeding when a given attack is detected.

3. The "System of Systems" Challenge

Many modern cyber-physical-human systems are composed of hundreds or even thousands of individual software systems. Each of these systems can be prone to failure or vulnerable to attacks in a myriad of diverse ways, but a focus solely on individual components is insufficient. In particular, interdependencies between systems must also be accounted for. Digital twins and other simulation tools can be used to model the behavior of interconnected systems, including human roles. These tools can help identify failure modes and the cascading effects of disruptions across different sectors. In addition, standards for testing resilience across multiple domains (e.g., physical, cyber, human) must be established. Recommendations for continuous testing and verification should be included in the national plan, especially as systems evolve and new AI-driven capabilities are added.

4. Incentivize Resilience and Resilience Research

A key difficulty in establishing a resilient digital ecosystem is the need for consistent enforcement. These systems are inherently interconnected, which poses a challenge to both prevention and isolation of breaches. Without universal enforcement of cybersecurity and resilience standards, a system can be left as vulnerable as the weakest link. Further, the stakeholders in the digital economy are diverse; simple government fiat is not sufficient to ensure compliance, absent a strong and efficient enforcement mechanism. Participants in the digital economy need an incentive to take these threats and challenges seriously.

There is an obvious solution to this issue. By making participants in the digital economy liable for damages caused by their security and resilience negligence, the government can ensure standards are immediately taken seriously. The potential damages for failure are massive, and to balance the risk of a breach or problem, economic actors would need to carry insurance against failures. In turn, insurance companies would promulgate security and resilience standards in line with the expectations of regulators. This system would work similarly to car, home, and malpractice insurance, where it is not direct government oversight driving compliance, but a self-interested insurance industry protecting against the risk of economic liability.

Additionally, efforts to holistically evaluate and overhaul cyber-physical systems will be expensive and require dedicated long-term funding to continue to test and evaluate systems as threats and attack methods evolve. We recommend funding fewer projects that are threat-agnostic and can be utilized to improve many cyber-physical systems. This can include projects like creating robust datasets for modeling and testing resilience, similar to successful initiatives in medical research. Funding threat-agnostic approaches can maximize flexibility in responding to emerging threats, including those that are unanticipated. The national plan should also recommend utilizing existing research centers and federal laboratories to spearhead resilience testing and experimentation, with a particular focus on critical infrastructure. Federally funded projects which aim to develop cyber-physical systems should have requirements for developing robust systems. These projects could also require the development of maintenance and testing plans to implement following the deployment of these systems.

5. Collaborate with Researchers

The PCAST report references the need to establish minimum operating capacities for critical systems to maintain essential functions during crises. This report mentions several government agencies which should assist in carrying out this task, such as NIST, however, it does not mention the important role that researchers will play. Engaging the research community in defining these capabilities would ensure they are realistic and robust.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Dennis S. Bernstein

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Research Directions for Resilient Control of Cyber-Physical Systems

Dennis S. Bernstein
Department of Aerospace Engineering
The University of Michigan
Ann Arbor MI, 48109-2140
[redacted]

October 7, 2024

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

Feedback control and uncertainty

Feedback control mitigates uncertainty. As an autonomous car drives or an autonomous aircraft flies, the control system uses data from sensors to compensate for myriad unknown effects that would otherwise prevent successful operation. Feedback control is essential for a vast number of technologies, but its presence and function are largely invisible.

Despite innumerable successes, many control applications are beyond the reach of existing techniques. These applications involve dynamics that change in unknown and unpredictable ways. Examples include the electric power grid, which changes as loads and sources are connected and disconnected; autonomous vehicles on roads whose coefficient of friction changes due to weather; and aircraft whose dynamics change due to payloads and damage. *In these and many other applications, it is extremely difficult to ensure that the control system will be resilient to unknown and unpredictable changes.*

Research Direction #1: Fast, targeted learning.

Machine-learning techniques use data-intensive methods to model the dynamics of a system, and the identified model is used to synthesize a feedback controller. Unfortunately, pre-training based on a system model produces a controller that is tuned for the model, which is invariably an imperfect representation of the true system. In addition, *machine-learning techniques are ineffective in the face of unknown and unpredictable changes, especially changes that have not occurred before and therefore cannot be anticipated by any amount of prior training.*

To address these challenges, resilient control of cyber-physical systems requires methods that can learn rapidly as the system changes. This can be done by *fast, targeted learning*, which rapidly updates system details that are crucial for closed-loop performance. Fast, targeted learning is an essential technology for resilient control of cyber-physical systems.

Research Direction #2: Adaptively cooperative decentralized control.

Systems such as the electric power grid are spatially distributed. This means that sensor data may not be available to a central processor. Instead, separate controllers must operate without direct communication. By operating in a decentralized fashion, the only "communication" among controllers is through the locally observed behavior of the system. To ensure resilient operation, decentralized controllers must learn to cooperate as the global system changes in unknown and unpredictable ways. Adaptively cooperative decentralized control is an essential technology for resilient control of cyber-physical systems.

Research Direction #3: Scientifically meaningful control experiments.

The foundation of control research is built on mathematics, as it should be. In cyber-physical systems, however, control algorithms interact with physical systems, and thus sensor data are generated by the full complexity of the physical system. Since no model can fully capture this complexity, *the challenge is to ensure that a cyber-physical system is resilient to unknown and unmodelable effects.* The ubiquity of control technology suggests that real-world control systems are inherently resilient. Many are, but the need for resilient control of cyber-physical systems demands a new paradigm for assessing the challenges of interacting with complex physical systems. The missing paradigm is *control experimentation.*

Physical experiments are essential for advancing all branches of science, but this has not been the case for feedback control. The lack of a culture of feedback-control experiments reflects the fact that the goal of feedback control is not to attain a deeper understanding of physical reality, but rather is to mitigate uncertainty in that understanding. Feedback control is thus unique among all branches of science.

It is tempting to view a control "experiment" as a physical construction whose features resemble the hypotheses of a theorem, in which case it may be expected that the conclusions of the theorem (such as stability and robustness) are ensured. The importance of neglected effects, however, can only be assessed by trial and error, engineering judgment, and experience. Unfortunately, learning from failure is often expensive.

A fundamental understanding of scientifically meaningful control experiments is needed. This question is deep and nontrivial, especially since the goal of feedback control is not to understand reality, but rather is to mitigate uncertainty. How does one ensure that feedback control mitigates uncertainty when the full complexity of physical reality cannot be fully known? These and related questions are not philosophical; rather, they get to the heart of resilient control of cyber-physical systems [1].

[1] Bernstein, D. S., "Facing Future Challenges in Feedback Control of Aerospace Systems Through Scientific Experimentation," *AIAA J. Guid. Contr.*, Vol. 45, 2022, pp. 2202–2210.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Dr. Peter A. Beling

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

RFI Response: Cyber-Physical Systems Resilience R&D Plan

Methods in Support of Secure Cyber Resilient Engineering

October 25, 2024

Dr. Peter A. Beling
Virginia Tech National Security Institute

“This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.”

1. **Introduction**

This response presents a **systems-centric approach to cyber-physical system (CPS) resilience**, aligned with the principles of **Secure Cyber Resilient Engineering (SCRE)**. SCRE focuses on integrating cybersecurity and resilience into system engineering activities across the life cycle. Core methodologies and decision support tools for SCRE were developed for the Office of the Undersecretary for Defense, Research and Engineering (OUSD (R&E)) by researchers at Virginia Tech and Stevens Institute of Technology under the auspices of the Systems Engineering Research Center (SERC), a DoD funded University Affiliated Research Center (UARC) [1-8].

Our approach has as its goal the **design of adaptive systems** capable of maintaining operational integrity under adverse conditions. Through selective engineering of **resilience mechanisms**, we aim to preserve CPS mission capability in the face of evolving and unpredictable threats. Our approach relies on a **loss-driven and threat-agnostic** framework for system modeling and risk characterization. The approach also leverages SCRE's focus on **multi-criteria trade-offs and behavior prediction**, providing decision-makers with practical tools to assess and enhance resilience.

2. Secure Cyber Resilient Engineering

The proposed framework combines **system-theoretic modeling, behavior prediction, and dynamic control strategies** to enhance the ability of CPS to preserve essential mission function under adversities created by adversaries, equipment failure, environmental factors, or other causes.

Resilience is treated as a **control problem**, with the focus on maintaining **safe and adaptive control actions** under normal and degraded conditions. Predictive models, based on system behavior, are used to anticipate disruptions and trigger **automated recovery actions**. This predictive capacity aims to provide:

- **Dynamic Adaptation**: The system can **adjust control actions** in real time to maintain operational integrity.
- **Proactive Intervention**: Through real-time monitoring, the framework **identifies early indicators** of failure and preemptively initiates mitigation strategies.

2.1. Systems-Theoretic Process Analysis

At the core of this framework is **Systems-Theoretic Process Analysis (STPA)**, a modeling technique that extends traditional hazard analysis by focusing on how **control structures and unsafe interactions** can contribute to system losses [9]. Unlike conventional reliability models, which concentrate on individual component failures, STPA provides a holistic view of **system control dynamics**, focusing on:

- **Unacceptable Losses and Hazardous States**: The framework identifies critical outcomes to avoid, ensuring that system behavior aligns with mission objectives even under degraded conditions.
- **Control Structures and Interactions**: Control structures within the CPS are modeled to identify how actions (or lack thereof) can trigger hazardous states or lead to cascading failures.
- **Causal Pathways**: STPA traces the interactions between system components, revealing how unsafe control actions can result from human error, software flaws, or unforeseen interactions.

This **top-down, hazard-driven approach** aims to integrate safety, security, and resilience into the design from the outset. The approach also uses Systems Theoretic Process Assessment for Security (STPA-Sec), an extension of STPA with an emphasis on adversarial environments and the modeling of security controls.

The following section describes how SCRE employs STPA to drive the engineering of resilience.

2.2. The SCRE Approach to Threat-Agnostic Resilience by Design

Mission Aware is an approach to cyber resilience-by-design in which the system is engineered to include one or more resilience mechanisms. A basic design pattern for a resilience mechanism would feature processes, called sentinels, that monitor for the symptoms of loss of system functionality or mission capability. In the event of a detection, a sentinel will attempt to reconfigure the system by engaging alternate sets of hardware and software designed to permit continued operation despite the attack. Sentinel-based resilience finds most of its application in cyber-physical systems, such as vehicles and weapons systems, rather than in pure cyber and networking systems such as enterprise information technology systems.

There are many potential technical approaches to the fundamental resilience tasks of detection, mitigation, and recovery. Examples of concepts include:

- **Redundant and Diverse Systems:** Redundant components sourced from multiple suppliers mitigate risks associated with supply chain vulnerabilities and increase system robustness.
- **Configuration Hopping:** Physical and virtual control elements shift dynamically among redundant nodes, disrupting potential attack chains and reducing system predictability.
- **Voting Mechanisms and Graceful Degradation:** Redundant control systems use voting protocols to validate critical actions, while graceful degradation ensures continued functionality during partial failures.

See [11] for a broader discussion of design patterns for resilience mechanisms. All mechanisms share the characteristic that they come with costs in terms of money, complexity, or operational performance. The frameworks described below aim to address the problem of selecting where and how to engineering these mechanism into the system.

The *Framework for Operational Resilience in Engineering and System Test (FOREST)* is a process meta-model that provides a decomposition of operational resilience into the principal mechanisms, options, information flows, and decisions that arise as attacks and resilience responses play out in systems [3, 4]. The framework is composed of eight elements known as Testable Requirements Elicitation Elements (TREEs). The first TREE embodies the notion that there is active sensing to detect loss of function or abnormal behavior in the system. Next, the framework considers the task of isolating a detected incident and the use of diagnostic information as the basis for choosing resilience mode responses. From that point, FOREST expands to include consideration of operator response and supporting technology. For instance, would an operator have confidence in resilience solutions being employed, or does the system provide the operator with the ability to run tests or exercise control to help in gaining confidence in resilience modes of operation. Finally, the framework considers decision support and archiving to allow for post-event analysis and adaptation.

There is significant complexity to the TREEs, and many of them overlap intentionally and deal with issues at the intersections of technology, doctrine, and people. As their name implies, TREEs provide a view of resilience that supports the development of test plans, and associated measures and metrics, for both the technological and operational aspects of the system.

Cyber Resilient Requirements Methodology (CRRM) is a risk-based methodology for addressing cyber security during the design phase of a cyber-physical system [3, 4]. CRRM is intended for use by a multidisciplinary evaluation team reflecting knowledge of the systems operational context, the system design, the cyber threat, and the ability to effectively test:

- Systems Engineering (SE) Team: Responsible for managing the CRRM process and developing system designs and definitions that reflect requirements, objectives, constraints, and stakeholder concerns, and for ensuring the current system design, including resilience modes of operation, can be adequately tested.
- Blue Team: Composed of operationally-oriented members with experience using similar systems. The blue team is responsible for providing consequences and risks to the CRRM process.
- Red Team: Composed of cyber security experts and cyber-attack experts who will provide the likelihood of different attacks given the current system design and resilient solutions.
- Grey Team: Composed of system/operational test experts who will evaluate test and measurement approaches given the current system design and resilient solutions.

CRRM is an integration of the STPA-Sec and FOREST methodologies, based on a Mission Aware model-based systems engineering (MBSE) meta-model. CRRM helps stakeholders identify loss scenarios that are evaluated to determine remediation mechanisms which effectively minimize the loss using sentinel detection patterns and resilience architecture patterns. The architectural tradespace incorporates the set of sentinels and resilience modes which mitigate the most likely cyber-attacks which could lead to the highest priority mission losses that are within the programmatic constraints of development time and budget.

2.3. The Adversity Chain: Contrasting Prevention with Resilience

In the a canonical Cyber Kill Chain (see, e.g., [[12]) , the idea is to show the sequence of categories of activities that an adversary might follow in progressing toward culminating exploit, that we term a *loss scenario*. The loss scenario can be viewed as the point of final action or control on the part of the adversary. Prevention methods in conventional cybersecurity are designed to reduce the likelihood of a loss scenario being realized.

An alternate perspective, and one that is central to resilience, is to reason about how our system might operate given a loss scenario as a starting point. Following the concepts of STPA-sec, the Adversity Chain models a sequence of actions and system state transitions, starting from the state of the loss scenario and possibly ending in a loss state, as defined by the mission and system owners.

These two chain models can be used to frame the problem of achieving cyber survivability. Prevention techniques and practices are used to break Cyber Kill Chains; that is, prevention aims to keep an adversary from progressing to the loss scenario. Often, assurance cases, consisting of formal proofs or structured arguments, are developed to give the program confidence that kill chains are adequately accounted for in the system design. The fundamental perspective of resilience is that not all loss scenarios will be covered by an assurance case. Loss scenarios can and will occur. The CRRM methodology was created to address these cases. CRRM provides a structured approach to identifying key loss scenarios and architecting resilience mechanisms that will prevent the corresponding Adversity Chains from reaching the loss state. We call this “breaking the Adversity Chain”.

2.4. Multi-Criteria Trade-Offs for Resilience Design

A key feature of the SCRE approach is its ability to connect hazard analysis with the engineering trade space. Within the context of the Mission Aware MBSE meta-model, simulation can be used to identify trade-offs between performance, cost, complexity, and resilience. In system acquisition, these trade-offs are critical to setting of technical requirements for resilience mechanisms.

A quantification of mission/system resilience can be derived by indirectly measuring the effectiveness of sentinel scenarios and associated resilience mechanisms in breaking the Adversity Chain. In this context, the behavior of the mission/system is defined through a set of state and activity specifications for each of the relevant system components, external actors, and environmental interactions. A mission profile is defined using a subjective probability distribution to specify the duration of an activity or state while probabilities are defined for action decision paths in the behavior specifications. The injection of adverse behavior, as defined by identified loss scenarios in CRRM/STPA, is accomplished using a test support system. The sentinel and associated scenarios provide mechanisms to vary the FOREST-based requirement parameters (e.g., sense time, resilience execution time) for associated loss scenarios to understand their effect on mission loss.

4. Current Research: Wind Energy and Critical Infrastructure

In our ongoing research, the Virginia Tech National Security Institute is applying the SCRE methodologies to two classes of use cases:

- **Wind Energy Farms**: Specifically, we are studying cyber resilience for offshore wind energy farms using a pilot site belonging to Dominion Energy. This effort includes collaboration with Stevens Institute of Technology and Old Dominion University in the context of a Center for Offshore Wind Energy, which has a security and resilience focus.
- **Critical Infrastructure Systems**: We are developing a model-based test-bed to represent another a broader class of distributed energy systems.

In the context of the use case, we are studying how SCRE could be enhanced through incorporation of ideas from related methodologies. SCRE provides a foundation for secure system engineering, while **Operational Technology Assurance (OTA)** principles address resilience within **nuclear and energy operations**. OTA, as defined by the DOE's National Nuclear Security Administration (NNSA), focuses on identifying and mitigating **cyber risks specific to OT environments**, such as safety systems, additive manufacturing, and processes that control physical operations.

The OTA framework emphasizes **continuous monitoring and operational assurance** to address risks across the entire system lifecycle, including supply chains and logistics. The integration of **OTA methodologies** with SCRE could enhance the use of **real-time control and physical system adaptation** in resilience engineering.

5. Recommendations for the National Cyber-Physical Systems Resilience R&D Plan

1. **Support the Integration of SCRE and related methodologies such as OTA in the systems engineering process**: Encourage research that aligns **engineering and resilience methodologies**, ensuring secure and adaptable systems.
2. **Develop Tools for Multi-Criteria Resilience Assessment**: Invest in tools that **balance performance, cost, and security** across cyber-physical systems and environments.
3. **Promote Cross-Sector Collaboration**: Engage stakeholders from both public and private sectors to ensure **scalable resilience strategies** aligned with real-world needs.
4. **Expand Early-Phase Threat Modeling and Verification**: Provide automated tools for **threat modeling and vulnerability assessment** during the system design phase to improve cost-efficiency.
5. **Invest in Workforce Development**: Equip engineers and operators with **training in viewing resilience as an engineering topic**, ensuring the next generation is prepared to architect, design, and implement resilient systems.

5. Conclusion

This response outlines a **comprehensive framework for secure cyber-resilient engineering**, combining the strengths of **SCRE and OTA** to address both engineering and operational challenges. By integrating

real-time monitoring, dynamic control strategies, and adaptive recovery mechanisms, the framework ensures that CPSs maintain mission-critical operations in the face of evolving threats.
We look forward to contributing to the **National Cyber-Physical Systems Resilience R&D Strategic Plan** and participating in efforts to advance the resilience of critical infrastructure across the nation.

References

1. Beling, P., Horowitz, B., Fleming, C., Adams, S., Bakirtzis, G., Carter, B., ... & Simon, B. (2019). *Model-Based Engineering for Functional Risk Assessment and Design of Cyber-Resilient Systems*. University of Virginia, Charlottesville, United States.
2. Horowitz, B., Beling, P., Fleming, C., Adams, S., Carter, B., Sherburne, T., ... & Mead, N. R. (2018). *Cybersecurity Requirements Methodology*. Stevens Institute of Technology, Hoboken, United States.
3. McDermott, T., Clifford, M. M., Sherburne, T., Horowitz, B., & Beling, P. A. (2022). Framework for Operational Resilience in Engineering and System Test (FOREST) Part I: Methodology – Responding to "Security as a Functional Requirement." *INSIGHT, 25*(2), 30–37. https://doi.org/10.xxxx/insight.2022.25.2.30
4. McDermott, T., Clifford, M. M., Sherburne, T., Horowitz, B., & Beling, P. A. (2022). Framework for Operational Resilience in Engineering and System Test (FOREST) Part II: Case Study – Responding to "Security as a Functional Requirement." *INSIGHT, 25*(2), 38–43. https://doi.org/10.xxxx/insight.2022.25.2.38
5. Beling, P., Horowitz, B., Fleming, C., Adams, S., Bakirtzis, G., Carter, B., Sherburne, T., Elks, C., Collins, A., & Simon, B. (2021). *Developmental Test and Evaluation (DTE&A) and Cyberattack Resilient Systems*. Stevens Institute of Technology, Hoboken, United States.
6. Carter, B., Adams, S., Bakirtzis, G., Sherburne, T., Beling, P., Horowitz, B., & Fleming, C. (2019). A Preliminary Design-Phase Security Methodology for Cyber–Physical Systems. *Systems, 7*(2), 21. https://doi.org/10.xxxx/systems.2019.7.2.21
7. Fleming, C. H., Elks, C., Bakirtzis, G., Adams, S., Carter, B., Beling, P., & Horowitz, B. (2021). Cyberphysical Security Through Resiliency: A Systems-Centric Approach. *Computer, 54*(6), 36–45. https://doi.org/10.xxxx/computer.2021.54.6.36
8. Horowitz, B., Beling, P., Fleming, C., Adams, S., Carter, B., Vemuru, K., Elks, C., Bakker, T., Cios, K., Bakirtzis, G., & Collins, A. (2017). *Security Engineering FY17 Systems-Aware Cybersecurity*. Stevens Institute of Technology, Hoboken, United States.
9. Leveson, N. G. (2004). A systems-theoretic Approach toSafety in Software-intensive Systems. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 66-86.
10. Fleming, C. (2023). Introduction to STPA-Sec. *Systems Engineering for the Digital Age: Practitioner Perspectives*, 489-505.
11. Beling, P. A., Sherburne, T., & Horowitz, B. (2023). Case Study C: Sentinels for Cyber Resilience. In *Autonomous Intelligent Cyber Defense Agent (AICA) A Comprehensive Guide* (pp. 425-445). Cham: Springer International Publishing.
12. Naik, N., Jenkins, P., Grace, P., & Song, J. (2022, October). Comparing attack models for it systems: Lockheed martin's cyber kill chain, mitre att&ck framework and diamond model. In *2022 IEEE International Symposium on Systems Engineering (ISSE)* (pp. 1-7). IEEE.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Fanxin Kong

Insup Lee

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Real-time and Adaptive Resilience for Cyber-Physical Systems[1]

Fanxin Kong, University of Notre Dame
Insup Lee, University of Pennsylvania

Overview. Cyber-Physical Systems (CPS) are pervasive in vital sectors such as transportation, energy, manufacturing, and so on. The disruption to these systems could have debilitating impacts on our society due to our critical reliance on their correct operations. CPS is vulnerable to a wide range of threats including cyber attacks, environment manipulation, and physical damages due to the tight coupling of cyber components and physical systems. These threats can severely disrupt vital sectors and cause catastrophic consequences. Hence, enhancing CPS resilience against those threats is essential.

Research Gaps. Although there are many research works on CPS resilience, several key challenges remain inadequately or barely addressed.

i) Unpredictability of threats. It is extremely difficult, if not impossible, to exactly know the occurrence of attacks beforehand. Two types of unpredictability need to be studied. One is spatial unpredictability, that is, what components in CPS will be compromised by an attack. The other is temporal unpredictability, that is, when an attack will occur. Effective CPS resilience solutions need to accommodate both the spatial and temporal unpredictability.

ii) Timing requirements. Untimely defense is just as damaging. Response to attacks, both cyber and physical attacks, need to be carried out before irreparable consequences occur. For example, a CPS maybe not recoverable if an attack is detected after the system crashes.

Research Needs. The needed research aims at the target of ensuring CPS safety while maximizing the performance in presence of cyber, environmental, and physical attacks. The challenges above are in particular needed to be addressed.

i) Adaptive resilience. Extensive studies have been carried out for attack detection in CPS. However, they either i) focus on one vector (i.e., sensors, controllers, or actuators) in the control loop and assume other vectors are uncompromised, or ii) just report the occurrence of an attack without pointing out which vector is under attack. For case i), the assumption is not realistic due to the unpredictability of threats, i.e., every vector can be corrupted. We need to tell which vector and which dimensions in a vector are under attack in the first place, instead of not dealing with this by assuming certain vectors are intact. For case ii), binary detection results without pointing out compromised vectors provide little information useful for the reaction to the attack (e.g., recovery).

To overcome these limitations, a holistic and threat-agnostic resilience approach is needed for CPS. One approach is to consider the control loop as a whole from the beginning and can accommodate different compromised vectors by adapting its reaction accordingly. Distinguishing attacks on different vectors is non-trivial. For example, attacks on sensors and actuators may both manifest as interference signals to the controller and drift the physical system. The solution needs to include i) threat-agnostic detection to find if a CPS is under attack, ii) spatial-temporal diagnosis to determine which vectors and dimensions are compromised and when that started, and iii) multi-mode recovery that can restore a CPS and accommodate different compromised vectors. Multi-mode recovery design needs to associate each mode to a vector and the corresponding recovery

[1] This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution

approach, and then adopt the right mode according to the specific compromised vector online.

ii) Real-time resilience. As mentioned above, a CPS needs to be safely recovered while maintaining acceptable performance. This requires sufficient timeliness for both recovery and detection. That is, a timely recovery requires that detection makes early enough decisions to allow enough time for the subsequential response to keep a CPS from becoming unsafe. For example, if a detector raises an alarm too late, there may not be enough time for a system to be recovered to a safe state even if the detection result is accurate.

Existing studies, however, address attack detection and recovery separately, instead of coordinating the two parties. On the one hand, existing detection methods mainly seek to increase the accuracy of the detection result, while overlooking whether a system can be recovered to a safe state after the detection. On the other hand, existing recovery methods usually assume a detector already in place and thus are subject to the accuracy of the detection. That is, these recovery methods do not consider how the recovery should affect the detection. Therefore, we posit the importance of coordinating the detection, diagnosis, and recovery for a real-time and safe defense against attacks. The solution needs to dynamically adjust the sensitivity and timeliness of detection and diagnosis to guarantee timely and safe recovery once an attack is detected.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Hongkai Chen

Shan Lin

Nicola Paoletti

Scott A. Smolka

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Specification, Verification, and Control for Resilient Cyber-Physical Systems:
Response to NSF RFI on CPS Resilience

Hongkai Chen[1], Shan Lin[2], Nicola Paoletti[3], and Scott A. Smolka[4]

[1]Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong
[2]Department of Electrical and Computer Engineering, Stony Brook University
[3]Department of Informatics, King's College London, UK
[4]Department of Computer Science, Stony Brook University

October 25, 2024

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

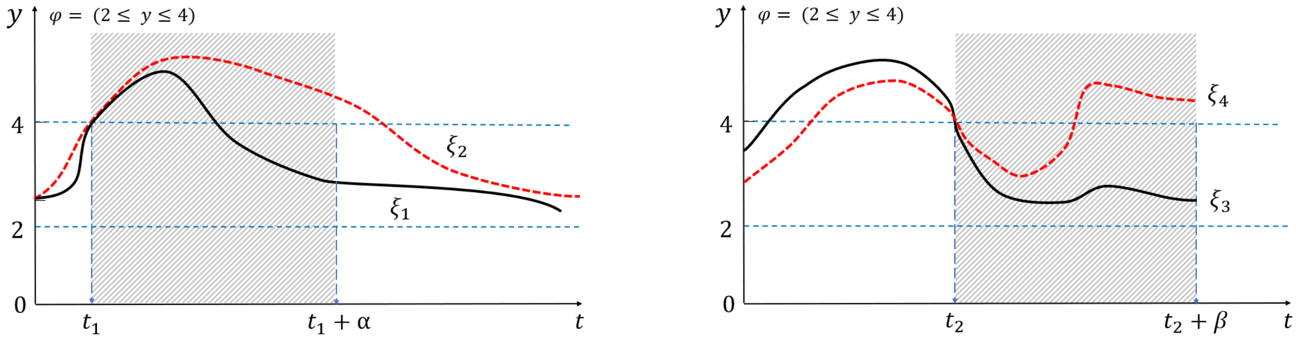
1 Introduction

Resiliency (*syn.* resilience) is defined as the ability to recover from or adjust easily to adversity or change [7]. Resiliency is of fundamental importance in Cyber-Physical Systems (CPS), which are expected to exhibit safety- or mission-critical behavior even in the presence of internal faults or external disturbances. Consider for example the *lane keeping* problem for autonomous vehicles (AVs), which requires a vehicle to stay within the marked boundaries of the lane it is driving in. The standard temporal-logic-based notion of safety is not ideally suited for specifying the AV's behavior when it comes to lane keeping. This is because AV technology is not perfect and driving conditions (e.g., being crowded by a neighboring vehicle) and other external disturbances may require occasional or even intermittent violations of lane keeping. Rather, the AV should behave resiliently in the presence of a lane violation, recovering from the violation in a timely fashion (*recoverability*), and avoiding future lane departures for as long as possible (*durability*). Unfortunately, there is no widely agreed upon notion of resiliency within the CPS community, despite several efforts to settle the issue,

Moreover, reasoning about resiliency in CPS and creating resilient systems are considered to be open problems in the recently published *Control for Societal-scale Challenges: Road Map 2030* [2]. To address this state of affairs—and to simultaneously point out how well-aligned our research on CPS Resilience is with the *NSF RFI*, we have recently made two foundational contributions to the theory and practice of CPS resiliency:

Contribution 1: In [4], we developed an STL [6]-based framework for specifying and reasoning about resiliency in CPS in which resiliency has a syntactic characterization in the form of an *STL-based Resiliency Specification* (SRS). Given an STL formula φ, time bounds α and β, the SRS of φ, denoted $R_{\alpha,\beta}(\varphi)$, requires that recovery from a violation of φ occur within time α (*recoverability*), and subsequently that φ be maintained for duration β (*durability*). These R-expressions, which are atoms in our SRS logic, can be combined using STL operators, allowing one to express composite resiliency specifications, e.g., multiple SRSs must hold simultaneously, or the system must eventually be resilient. Further details about this work are provided in the Prior Work section of this proposal (Section 2.1).

Contribution 2: In [5], we present *ResilienC*, a framework for resilient control of Cyber-Physical Systems subject to STL-based requirements. We define the *Resilient STL Control Problem* as one of *multi-objective optimization*, where the recoverability and durability of the desired STL specification are maximized. When neither objective is prioritized over the other, the solution to the problem is a set of *Pareto-optimal* system trajectories. We present a precise solution method to the problem using a mixed-integer linear programming encoding and an *a posteriori* ϵ-constraint approach for efficiently retrieving the complete set of optimally resilient solutions. In ResilienC, at each time-step, the optimal control action selected from the set of Pareto-optimal



(a) Signal ξ_1 satisfies recoverability.

(b) Signal ξ_3 satisfies durability.

Figure 1: Resilience w.r.t. STL formula $\varphi = (2 \leq y \leq 4)$

solutions by a *Decision Maker* strategy realizes a form of *Model Predictive Control*. Further details about ResilienC are available in Section 2.2.

2 Our Prior Work on CPS Resilience

In this section, we provide the reader with further details regarding our two contributions to CPS Resilience outlined in Section 1.

2.1 An STL-based Formulation of Resiliency

In a recent paper [4],[1] we formulate CPS resiliency using *STL-based Resiliency Specifications* (SRSs). Given an arbitrary STL formula φ (STL stands for Signal Temporal Logic [6]), time bounds α and β, the SRS of φ, $R_{\alpha,\beta}(\varphi)$, is the STL formula $\neg\varphi\mathbf{U}_{[0,\alpha]}\mathbf{G}_{[0,\beta)}\varphi$, specifying that recovery from a violation of φ occur within time α, and subsequently φ be maintained for duration at least β. The SRS of φ captures the requirement that a system quickly recovers from a violation of φ (*recoverability*) and then satisfy φ for an extended period of time (*durability*). $R_{\alpha,\beta}(\varphi)$ expressions, which are atoms in our SRS logic, can be inductively combined using STL operators, allowing one to express composite resiliency specifications; e.g., multiple SRSs must hold simultaneously ($R_{\alpha_1,\beta_1}(\varphi_1) \wedge R_{\alpha_2,\beta_2}(\varphi_2)$), and the system must eventually be resilient ($\mathbf{F}_I R_{\alpha,\beta}(\varphi)$).

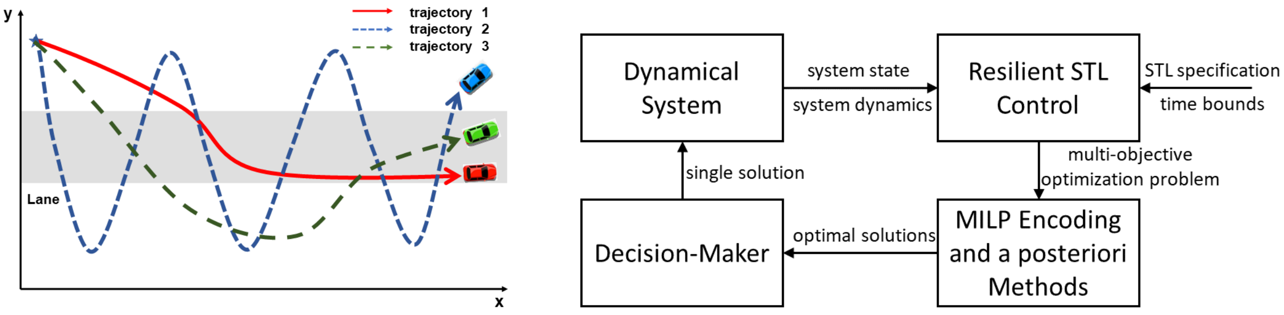
Consider for example the STL specification $\varphi = (2 \leq y \leq 4)$, where y is a signal. In Figure 1(a), signals ξ_1 and ξ_2 violate φ at time t_1. Given recovery deadline α, we see that only ξ_1 satisfies recoverability of φ w.r.t. α because φ becomes true before $t_1 + \alpha$. In the case of ξ_2, φ becomes true only after $t_1 + \alpha$. In Figure 1(b), signals ξ_3 and ξ_4 recover to satisfy φ at time t_2. Given durability bound β, we observe that only ξ_3 is durable w.r.t. β.

We can quantify the level of resiliency exhibited by a system trajectory by defining a so-called "quantitative semantics" for SRSs in the form of a *Resiliency Satisfaction Value* (ReSV) function r. Our semantics for $R_{\alpha,\beta}(\varphi)$ atoms is a singleton set of the form $\{(rec, dur)\}$, where *rec* quantifies how early before bound α recovery occurs, and *dur* indicates for how long after bound β property φ is maintained. The r-value for a composite SRS formula results in a set of non-dominated recoverability-durability pairs, given that the ReSVs of subformulas might not be directly comparable (e.g., one subformula has superior durability but worse recoverability than another). To the best of our knowledge, this is the first *multi-dimensional* quantitative semantics for STL.

2.2 Resilient Control

The formulation of resilience presented in Section 2.1 naturally suggests the problem of *resilient STL control*: find an optimal control strategy that maximizes the system's resilience in terms of recoverability and durability. These two objectives are often at odds with each other. For example, in the *lane-keeping problem* (see Figure 2(a)), the initial location of the vehicle, marked by a star, is outside the lane, thereby violating a lane-keeping specification φ. Trajectory 1 is the slowest to recover from its violation of φ, resulting in the worst recoverability among the three trajectories shown in the figure. However, the vehicle subsequently maintains φ until the end of the trajectory, resulting in the best durability. In trajectory 2, the vehicle aggressively steers back into the lane whenever φ is violated; so it exhibits the best recoverability. It cannot, however, maintain φ post-satisfaction due

[1] Received the Best Paper Award at the 2022 FORMATS conference.



(a) Illustration of resilient optimal solutions.

(b) The ResilienC architecture.

Figure 2: Illustration of resilient trajectories and Overview of ResilienC control framework.

to overshooting, and thus has the worst durability. The behavior of trajectory 3 lies in-between trajectories 1 and 2. The three trajectories are *mutually non-dominated* and, hence, equally resilient.

To address the resilient STL control problem, we have developed the *ResilienC* control framework, which formulates this problem as one of multi-objective optimization, designed to maximize both the recoverability and durability of the CPS under consideration [5]. Unlike existing techniques for STL-based control [8, 9] which focus on optimizing a single objective (e.g., spatial robustness in [8] and time robustness in [9]) and thus produce a single solution, our method results in a set of *non-dominated, aka Pareto-optimal, solutions.* Such a method is also called *a posteriori* as it avoids making any *a priori* assumptions about the relative importance of the two objectives. We realize an MPC scheme with our approach by deploying a *Decision Maker* strategy that at each time-step, selects the next optimal control action from among the set of Pareto-optimal solutions for execution by the plant. See Figure 2(b) for an overview of the ResilienC framework.

3 Proposed Work

The NSF RFI on CPS Resilience is very thought provoking and leads us to consider the following research activities as "Proposed Work". In particular, we put forth the *ReSCVe CPS* framework for Multi-Dimensional Resiliency Specification, Control, and Verification of CPS. The *ReSCVe CPS* effort (pronounced "Rescue CPS") will be centered around the following research tasks (please also see Figure 3).

Task 1: Specifying and reasoning about both *spatial* and *temporal* resiliency in CPS. A logical framework for reasoning about CPS resilience in the spatial-temporal domain will be developed. Resilient CPS subjected to external disturbances are expected to recover from such events and subsequently maintain desired behavior that may be subject to spatial and temporal requirements. Case studies will be used to demonstrate the expressiveness and utility of spatial-temporal resiliency specifications.

Task 2: Accurate and efficient predictive monitoring of CPS resiliency properties. We will devise methods to reliably predict the recoverability and durability of CPS at runtime, deriving prediction intervals with probabilistic coverage guarantees.

Task 3: Resilient control under uncertainty using MPC and Reinforcement learning. We will investigate resilient control techniques for CPS subject to two kinds of uncertainty: stochasticity (e.g., random disturbances) and non-determinism (when uncertainty cannot be quantified).

Task 4: Investigate the *Inherently resilient vs. Engineered-to-be-resilient systems* dichotomy. We believe that much is to be gained from understanding the mechanisms deployed in inherently resilient CPS (especially those occurring in nature) and applying these techniques to engineered CPS. For example, the human pancreas is an inherently resilient system for regulating human blood glucose levels, while the artificial pancreas is an engineered-to-be-resilient version of this digestive system organ.

Task 5: Perform an extensive performance evaluation of the ReSCVe CPS framework centered around three significant Case Studies. Case studies will involve demonstrating the resilience of the following CPS: (1) the Artificial Pancreas [3] (AP) subject to hypo- and hyper-glycemia; (2) Networked Microgrids [1] (NM) in the presence of varying loads, disconnection of distributed energy resources, and islanding from the main power grid; and (3) collective autonomous agents in *flock formation* subject to various external disturbances, including (artificial) wind, obstacles, and highly agile predator agents.

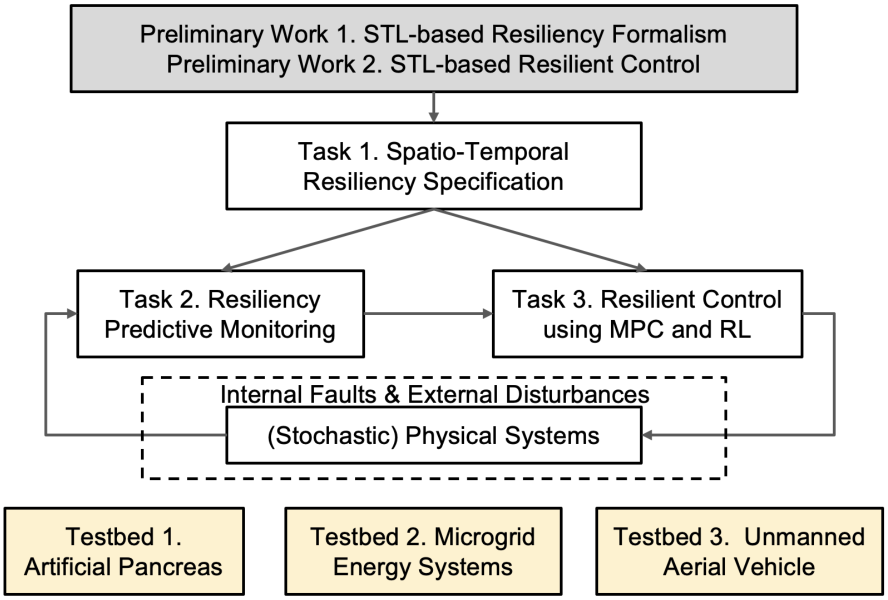

Preliminary Work 1. STL-based Resiliency Formalism
Preliminary Work 2. STL-based Resilient Control
Task 1. Spatio-Temporal Resiliency Specification
Task 2. Resiliency Predictive Monitoring
Task 3. Resilient Control using MPC and RL
Internal Faults & External Disturbances
(Stochastic) Physical Systems
Testbed 1. Artificial Pancreas
Testbed 2. Microgrid Energy Systems
Testbed 3. Unmanned Aerial Vehicle

Figure 3: Architectural overview of proposed *ReSCVe CPS* framework.

4 Conclusion

We have presented/highlighted our recent work on CPS Resilience, especially when it come to formal specification, verification, and control, and outlined proposed work in the area focused around the ReSCVe CPS framework. Our approach to CPS resilience enjoys the following benefits.

- Our logic-based framework for CPS resilience is foundational because it can be applied to *runtime monitoring*, *online/MPC-based control* (resilient-by-intervention), and *synthesis of fixed policies/controllers/system parameters* (which can be seen as resilient-by-design). Each of these approaches can be traditional/mechanistic or data-driven.

- To make our approach resilient to unseen threats/adversaries, we can solve all of the above problems in a robust (minmax) or uncertainty-aware (or risk-aware) manner.

- Our current resilient control approach can be made more efficient by considering "offline" policies or solving the single-objective problem (rather than bi-objective, which may be much more expensive) with a single recoverability-durability tradeoff value (which could change/adapt at runtime).

We would welcome the opportunity to pursue our CPS Resilience research goals within the confines of the NSF CPS program, and plan to submit a proposal to the same.

References

[1] Alam, M. N., Chakrabarti, S., and Ghosh, A. Networked microgrids: State-of-the-art and future perspectives. *IEEE Transactions on Industrial Informatics 15*, 3 (Mar. 2019).

[2] Alleyne, A., Allgöwer, F., Ames, A., et al. *Control for Societal-scale Challenges: Road Map 2030*. IEEE Control Systems Society, May 2023.

[3] Boughton, C. K., and Hovorka, R. Advances in artificial pancreas systems. *Science translational medicine 11*, 484 (2019), eaaw4949.

[4] Chen, H., Lin, S., Smolka, S. A., and Paoletti, N. An STL-based formulation of resilience in cyber-physical systems. In *Proceedings of FORMATS 2022, 20th International Conference on Formal Modeling and Analysis of Timed Systems, Best Paper Award* (2022).

[5] Chen, H., Smolka, S. A., Paoletti, N., and Lin, S. An STL-based approach to resilient control for cyber-physical systems. In *HSCC '23: Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control* (May 2023).

[6] Maler, O., and Nickovic, D. Monitoring temporal properties of continuous signals. In *Proceedings of Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*. Springer, Sept. 2004, pp. 152–166.

[7] Merriam-Webster Dictionary. https://www.merriam-webster.com/dictionary/resiliency.

[8] Raman, V., Donzé, A., Maasoumy, M., Murray, R. M., Sangiovanni-Vincentelli, A., and Seshia, S. A. Model predictive control with signal temporal logic specifications. In *Proceedings of the IEEE Conference on Decision and Control* (Dec. 2014), IEEE, pp. 81–87.

[9] Rodionova, A., Lindemann, L., Morari, M., and Pappas, G. J. Time-robust control for STL specifications. In *Proceedings of IEEE Conference on Decision and Control* (Dec. 2021), IEEE.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

J. Sukarno Mertoguno

Gregory Briskin

Jason H. Li

Kyung Kwak

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Chapter 1
Physics, Cyber-systems & Machine Learning

J. Sukarno Mertoguno, Gregory Briskin, Jason H. Li and Kyung Kwak

Abstract Cyber-physical systems (CPS) are used in various safety-critical domains such as robotics, industrial manufacturing systems, and power systems. Faults and cyber attacks have been shown to cause safety violations, which can damage these systems and endanger human lives. The past decade has seen the proliferation of research efforts related to security and resilience in cyber physical systems, with an abundance of publications, workshops, and even media attention. More recently, artificial intelligence (AI) and machine learning (ML) have been reinvigorated and become the topic of paramount attention across the research community, mass media and society, highlighted by trademark successes in the fields of gaming, video, audio, and language translation. While it seems natural to apply AI/ML in CPS security and resilience, the authors would like to share some lessons learned and future directions as cautionary notes, which include: (1) the critical importance of physics and the physical world (P); (2) various means and effects that are introduced by cyber (C); (3) interactions and ramifications of P and C in a system (S); (4) system model and control in CPS; (5) enhanced robustness of control and autonomy in CPS by AI/ML; (6) pitfalls and appropriate positioning of AI/ML in CPS security and resilience; and (7) some challenges and opportunities for research and development.

J. Sukarno Mertoguno
School of Cybersecurity & Privacy, Georgia Institu[redacted]

Gregory Briskin, Jason H. Li, and Kyung Kwak
Trusted Science and Technology Inc.[redacted]

1.1 Introduction

Cyber physical systems (CPS) underlie many critical infrastructures and are prevalent across a wide range of areas including the electrical grid, factory production pipeline, machinery control, vehicular control, internet-of-things (IOT) devices, and commodity toy drones, just to name a few. By its nature, a CPS straddles the continuous-time physical domain and the discrete-time digital or cyber domain. Cyber components (e.g., communication and computing) couple with physical components (e.g., sensors and actuators) to carry out the intended functions of the CPS.

Cyber physical systems are required to satisfy safety constraints in various application domains such as robotics, unmanned vehicles (e.g., aerial or ground), industrial manufacturing systems, and power systems. However, the once isolated system of computer-controlled machinery is now more exposed to the external world than ever, which renders ample opportunities of remote system disruption via cyber threats, in addition to the tradition threat of a physical component failing. Both may result in safety violations.

Current emphasis on cyber security of CPS is on securing the operational technology (OT) network. For example, National Institute of Standards and Technology (NIST) devoted its guidance for securing CPS, SP 800-82 Rev.3 [1], solely to network security with network segmentation as the primary recommended solution. However, network or communication is only one facet of CPS. While network is an important CPS and critical infrastructure component, over-emphasis on networking security will not be sufficient for defending the underlying CPS and its infrastructure against motivated and well-resourced adversaries. A recent article indicates that the assumption of malicious events entering the system solely via the external network (hence the need for network segmentation) has been invalidated [2]. The exploits discussed in the article did enter the system through the local (internal) network and propagated within the internal bus, avoiding the security protection provided by network segmentation. A holistic view and approach for defending CPS is needed.

1.2 Physical Domain and Cyber Domain

In CPS, the ultimate goal is for the overall system to be stable and function as intended. A *resilient CPS* is expected to physically operate properly and in a predictable and controllable manner under ever-present external and environmental disturbance as well as adversarial cyber exploits. The objectives and emphasis for CPS resilience are *physical* stability and functionality. Cyber components and systems in CPS are means toward the end of achieving CPS resilience. The stability of cyber systems by itself is not the primary objective.

A cyber physical system contains cyber components that interact with and control the behavior of the physical system operating in a physical environment. Generally speaking, the cyber controller periodically samples the operation (or mission) objective (e.g., the expected set value of speed), measures the actual values of physical

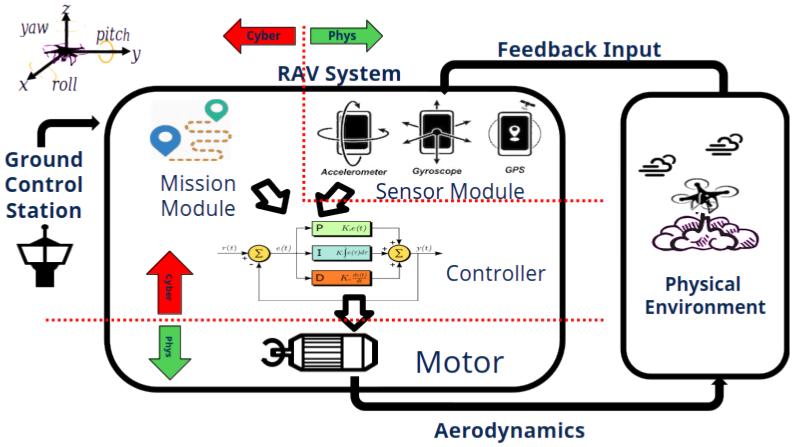


Fig. 1.1 Cyber Domain and Physical Domain in CPS, derived from MayDay [3]

variables via sensors (e.g., speed, altitude), contrasts measurements against the objective, and calculates the magnitude of control variables, which translates to direction and/or force to be asserted by the actuator onto the physical environment. Figure 1.1 shows an example of how the physical and cyber components interact in a particular cyber physical system – a robotic aerial vehicle (RAV) or drone.

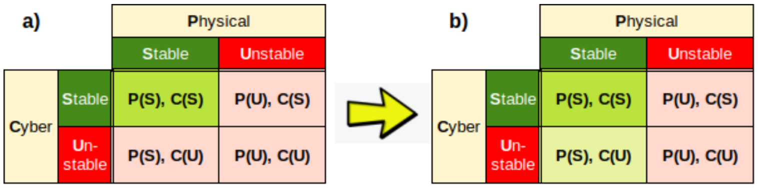
It is worthwhile to point out the differences of physical and cyber components and (sub)systems, and the potential opportunities they may offer for building resilient cyber physical systems. The physical platform and the subsystems operate in a physical environment at physical speed (and time), governed by the laws of physics. The mass and dynamics of a physical implementation define its moment of *inertia*, which in turn influences the *response time* of the physical subsystems and the platform. Any physical subsystem of a CPS must obey the laws of physics, and the physical systems invariably have inertia.

CPS physical and cyber components differ significantly in the scale of their response time. Physical and mechanical components have relatively large time scales (low frequency), in the order of milliseconds and seconds. A heavier object has larger inertia and hence lower frequency (see subsection 1.2.2.6 for detailed description of inertia). For example, a large tanker vessel takes minutes to change its direction.

The cyber components operate at cyber speed, typically multiple orders of magnitude faster than that of physical components. The scan period of a CPS controller (1-500 Hz) is usually about one or two orders of magnitude smaller (faster) than that of the physical/mechanical machinery [4]. The clock speed of the controller CPU (GHz) is generally five to seven orders of magnitude faster than the scan cycle. The physical micro-mechanical sensing mass within a micro-electro-mechanical systems (MEMS) inertia measurement unit (IMU) has resonance frequency measured in KHz, in the 10-30KHz range [5], still one or two orders of magnitude faster than the controller's scan cycle of a drone.

Traditionally CPS researchers have been focusing on achieving *cyber stability*, which generally provides physical stability within the designed region of operation. This is definitely a prudent design methodology (see the first quadrant in Figure 1.2a with P(S) and C(S) denoting physical and cyber stability, respectively).

However, a cyber physical system may have to operate in a physical environment with disturbance so large (e.g., strong wind gust or other physical impact) that makes the controller algorithms or other cyber components struggle to work while out of the designed region of operation. Extended Kalman Filter (EKF) and robust control algorithms usually work effectively to absorb and tolerate relatively small disturbance, but the physical subsystems and overall platform may fail to maintain physical stability under large disturbance. This is quite interesting: cyber components work as designed but the physical systems are unstable, see the second quadrant in Figure 1.2a with P(U) and C(S) denoting physical instability and cyber stability, respectively, which indicates that the traditional focus for cyber stability is not always sufficient or effective. The community starts to notice this important realization, and leaders start to investigate alternatives to designing resilient cyber physical systems, such as the DARPA LINC program [6] and the DARPA FIRE program [7].



a)

		Physical	
		Stable	Unstable
Cyber	Stable	P(S), C(S)	P(U), C(S)
	Un-stable	P(S), C(U)	P(U), C(U)

b)

		Physical	
		Stable	Unstable
Cyber	Stable	P(S), C(S)	P(U), C(S)
	Un-stable	P(S), C(U)	P(U), C(U)

Fig. 1.2 Desired CPS operation space: a) Cyber-centric, b) Physical-centric

Particularly, since we argue that the goal of CPS resilience is physical stability and not necessarily cyber stability, cyber components controlling physical components only need to be stable at the scale of the frequency and response time of the physical components. Focusing on physical stability, therefore, opens up another design space for CPS resilience as the fourth quadrant in Figure 1.2b, with P(S) and C(U) denoting physical stability and cyber instability, respectively. It is important to emphasize again that C(U) means being unstable but also unnoticeable by physical components, not being unstable all the time for obvious reasons. Empowered by physical inertia and the differences in response time for physical and cyber components, the fourth quadrant represents a previously less-explored design space for achieving CPS security and resilience. Subsection1.2.2.6 explores cyber-attack resilient CPS design within the fourth quadrant.

1.2.1 System Model and Control in CPS

Modeling is essential to every scientific and engineering enterprise. For both scientists and engineers, the "thing being modeled" (referred to as *target*) is typically an object, process, or system in the physical world. But it could also be another model as manifested in model refinement for formal verification. A "model" of a target is any description of the target that is not Kant's *thing-in-itself*. For example, mechanical engineers use Newton's laws as models for how a system will react to forces. Computer engineers model digital circuits as instruction set architectures (ISAs), programs as executions in an ISA, and applications as networks of program fragments [8]. Each of these models rests on a modeling paradigm. For example, a source code is a model of what a machine should do when it executes the program, but the source code is not what is actually run on a machine. The Java programming language, for example, is just such a modeling paradigm. Models abstract away details, and layers of models may be built on top of another. A CPS system consists of such layered models from hardware all the way up to applications it runs.

The *fidelity* of a model is the degree to which it emulates the target. When the target is a physical object, process, or system, model fidelity is never perfect. But as stated in reference [9], "essentially, all models are wrong, but some are useful". As highlighted in reference [8], in science the value of a model lies in how well its properties match those of the target, whereas in engineering the value of the target lies in how well its properties match those of the model. A scientist constructs models to help understand the target. An engineer constructs targets to emulate the properties of a model, since for an engineer a model represents a design and the target is the implementation. These two uses of models are complementary.

For CPS modeling and control, therefore, it is critical to always keep a clear mind in terms of the *thing*, the *model*, the *purpose* of the model, and the *interactions* between the thing and the model in either a science or engineering context. For example, simplicity and clarity of target semantics may dominate over accuracy and detail, and optimizing over a model does not necessarily bring about desired effects or benefits to the target.

Moreover, it is important to note that models (and analyses and controls over these models) have their inherent *region of operation*, a concept commonly known in each individual disciplines but unfortunately often ignored in real-world practices. This is particularly true in CPS where multi-layer models exist, and their interactions lack sufficient attention. The re-invigoration of AI/ML makes this awareness even more relevant, in terms of where and when AI/ML could help analyze and even take over some control of the cyber physical system without adversely affecting the physical or cyber operation stability. Special care must be taken to understand the boundary of each model, interactions among models, appropriate positioning of AI/ML models and algorithms, and anticipated and measurable effects in the physical world.

1.2.2 CPS-specific Cyber Security Challenges and Solutions

Traditional cyber defense for CPS has mainly focused on the level of human-machine interface (HMI) and security information and event management system (SIEM). This is largely due to the similarity to established cyber protections for hosts and networks, and the information technology (IT) mindset possessed by practitioners. However, this leaves the lower-levels of the cyber physical system vulnerable to attacks not common in a traditional IT environment.

Protection of low-level components and subsystems includes protecting the interconnect and computation or logic of the controllers. In general, cryptographic protections provide a way to disrupt potentially rogue modules from snooping at the bus. Although this is effective for protection, it might be considered unsuitable since the bus data is mainly useful only in real time when interpreted in context of the control model and physical situation. The overhead is simply too high for each involving module on the bus to conduct encryption and decryption constantly. In addition, compromises at the controller level, e.g., rogue control signals issued by the compromised controller, render encryption irrelevant (encrypting the rogue data does not help security or resilience), or even harmful since the attack traffic/attacker communication is protected by encryption.

Protecting the controllers themselves includes (and is not limited to) fault avoidance, fault tolerance, and model- or reference-based CPS security. Formal methods which attempt to reason over certain properties of an *implementation model* against a *specification model* is the dominant technique for fault avoidance. Considering the typical tractability and practicality of creating both models, using formal methods is an excellent approach for achieving CPS security and resilience via fault avoidance. But this would imply the complete redevelopment of the system (or at least the subsystem subject to formal methods) from scratch. This is very expensive for legacy systems which are prevalent in industry and military applications.

Fault tolerance is a complementary approach to fault avoidance. This method assumes that vulnerabilities exist in the controller code and strives to mitigate the effect of exploits by ensuring proper operations of the physical system part of CPS, even under successful cyber attacks, thus rendering CPS resilience. Fault tolerance methods often involve detection and recovery, including stateful component, subsystem or system level recovery.

Reference/model-based CPS security relies on the fact that a CPS, unlike general IT systems, is generally well constrained within its operation space and intended behaviors. The operations are periodic and predictable, and reference models for algorithms and the operating environment can be developed and used to detect discrepancy between the observed operation and models. Discrepancy beyond some tolerance threshold may indicate flaw, damage, disruption, or exploits.

1.2.2.1 Cyber Attacks against CPS and Critical Infrastructure

Our goal for CPS resilience is to have the physical systems behave properly regardless of fault or disruption (cyber or otherwise). In keeping with reality, we make no assumption that a system is devoid of bugs or vulnerabilities. Rather, we seek to enable a CPS to tolerate and live with existing bugs and vulnerabilities it may have.

We assume an Advanced Persistent Threat (APT)-like adversary, whose goal is to create maximum disruption, major damage, and difficult and lengthy recovery time. To defeat system protection & fault tolerance and to achieve maximum disruption & major damage, an adversary generally needs to subvert and affect many individual controllers (various systems components) simultaneously and in coordination. Uncoordinated one or two subversion and denial of service attacks are unlikely to cause major disruption or damage.

There are generally two methods to subvert or negatively-effect the behavior of a controller: (a) manipulate or inject malicious input to cause improper control output, or (b) hijack and own the controller via either rogue reprogramming command (from console) or malicious input that corrupt program execution, hijack the program control and own the controller. Note that cyber attacks that leak (confidential) information can be used to gather intelligence and help plan for an attack, but by itself cannot subvert the operational behavior of a cyber physical system.

To significantly influence a set of controllers of diverse functionalities and types, an adversary will need to inject many different inputs/signals in a coordinated manner, which is difficult to achieve in practice and often requires the adversary to own many controllers to perform coordinated, multiple signal injections. The most dangerous exploit is when devices/controllers were stealthily taken over one by one, and then upon triggering event(s), simultaneously act (in coordination) and disrupt the systems. Stealthily owning controllers are the prerequisite for APT-like coordinated attacks. An adversary can stealthily own a controller in several ways. One of them involves reprogramming (re-flashing) the controller itself, e.g., in the case of Stuxnet. To do this, the adversary will generally have to own either the maintenance laptop or the human machine interface (HMI) console, and issue malicious updates from the corrupted laptop or console. This risk can be reduced by requiring multi-factor authentication for firmware update/re-flashing. Another method for owning a controller would be to exploit a (software) vulnerability, and send/inject malicious inputs that will corrupt and take over the controller.

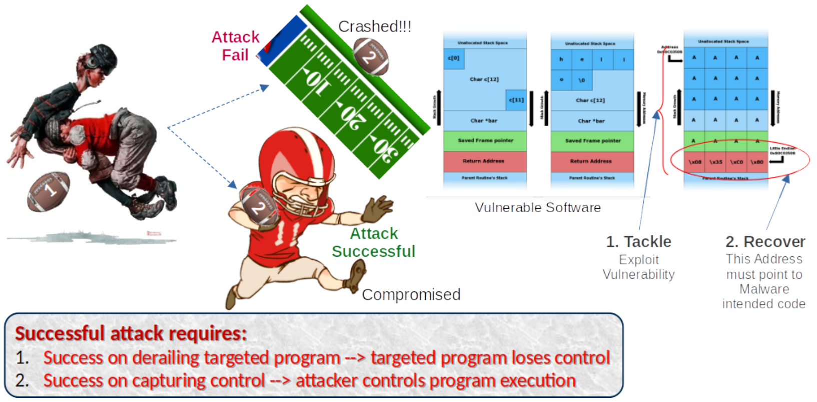
Byzantine fault tolerant++ (BFT++) is a family of cyber resilience methods that rely on the periodicity of CPS and the physical inertia to tolerate cyber attacks. The BFT++ family of CPS resilience prevents this particular class of methods for hijacking and owning the controller. Additionally, BFT++ is generic and agnostic to the particulars of malware or malicious inputs. Refer to subsection 1.2.2.6 for the detailed description of BFT++.

1.2.2.2 Anatomy of CPS/Controller Owning Cyber Exploits

Fault tolerance systems, such as byzantine fault tolerant (BFT) and quad redundant control (QRC), have been proven effective for safety critical systems. They rely on redundancy to detect and recover from faults, and essentially provide fault tolerance against natural disruption and random faults.

Cyber attacks present a new type of challenges. They can force faults in many components and subsystems simultaneously, which leads to a “common-mode failure” that traditional fault tolerance cannot effectively deal with. Worse, if the adversary is successful in compromising a component, there is no obvious fault signal to detect, and the controllers continue to actuate the system while compromised and under the control of an adversary. Attempts to deal with common-mode failures have been made through diversification, but the type of diversification must be appropriate to the class of causes of common-mode failures that the CPS owner wants to mitigate, and special care must be taken with respect to what, when and how much diversification is deployed depending on CPS and mission requirements.

The process of a cyber exploit involves two virtual stages: first, exploiting a flaw/vulnerability in the program’s code to alter its intended execution path, and second, taking control of the system to execute the attacker’s commands. This is analogous to a fumble in football, where an opposing team must not just cause a fault, but recover the ball to gain possession, as shown in Figure 1.3. A successful exploit will succeed in both stages, leading to compromised systems under attacker control. A condition when the first stage is successful but the second stage fails will generally manifest into a crash, due to corrupted (as opposed to compromised) cyber components and subsystems.



Successful attack requires:
1. Success on derailing targeted program --> targeted program loses control
2. Success on capturing control --> attacker controls program execution

Fig. 1.3 Two Virtual Stages of Cyber Exploit

1.2.2.3 Defense: White Listing and Operation Segmentation

Many CPS operational environments can be analyzed and segmented into several different modes of operation. Within each of these modes, the set of valid (allowed) operations can be whitelisted. Operation segmentation is analogous but orthogonal to network segmentation. While network segmentation limits exploits effects and propagation by limiting allowable communications, operation segmentation prevents CPS disruption by limiting incompatible and hazardous co-occurrence of operational commands/events.

For example, let us consider the cyber physical systems that control the operation of a ship, and for the purpose of illustration assume that the engineers decided to segment the ship operation into three modes: steaming mode, in-port mode, and maintenance mode. Maintenance mode is akin to the superuser mode in modern operating systems where (almost) all operations are allowed. For simplicity, let us consider three different operations: dropping anchors, brisk-steaming (above 5 knots), and re-flashing the controller. One can see that dropping-anchors and brisk-steaming are mutually exclusive. It will not be prudent to drop an anchor while briskly steaming, hence dropping anchors is not within the steaming mode whitelist, and brisk-steaming will not be in the in-port mode whitelist. Similarly, re-flashing a controller should only be performed in the maintenance mode.

Operation segmentation improves the operational safety of a cyber physical system. Separating the maintenance mode from other operation modes also enhances the system's cyber security posture by whitelisting out disallowed behaviors and requiring additional privilege for critical activities, such as re-flashing a controller. While it cannot completely prevent cyber attacks, operation segmentation erects barriers against various malicious activities that may otherwise readily perform once a foothold is obtained in a component or subsystem.

Operation segmentation focuses all the working aspects of the cyber physical systems onto the operators, which are responsible for the CPS operations, including approving and initiating CPS maintenance. This is a judicious method compatible with the principles of separation of duties and least privilege for building computer systems [10]. Current trends in modern cars, which are systems of cyber physical systems, however, are diverging from this philosophy. In the case of modern cars, it is the manufacturers who often initiate the system update, with or without the awareness of the operator (owner). There are both pro and con arguments that can be made for this context.

1.2.2.4 Defense: Reference Model Based CPS Security

Since cyber physical systems extensively communicate with their physical environment, system security relies not only on cyber security but also on securing the physical part of the system. This means the cyber layer, as well as the platform (including the physical) layer and their inter-dependency, must be considered together. For example, the platform layer covers the whole run-time environment containing artifacts like operating system and middleware, as well as the physical part of the system such as sensors and actuators, etc. Hence, as an entity that senses and interacts with the real world, a CPS could be exploited by an adversary and cause harmful impacts. Depending on the level of the attacker's access and capabilities, either or both sensing components and control software can be subjects of a compromise.

One of the most common security approaches for detection of attacks against control software is a comparison of true and faulty signals, thus necessitating trusted redundancies. For example, if an extra electronic control unit (ECU) hardware is retrofitted to the robotic vehicle with no access channel from the outside, it is shielded from the attacker, and hence can be trusted. Such CPS can still operate as intended with its original control software, while the control signal can also be used to enable comparison against the retrofitted ECU for attack detection and response. Instead of changing the original control system, an external piece of hardware can be used to monitor the given ECU with minimum modification to the original system. Independently implementing the CPS control and sensing logic software on the external hardware enables high-accuracy error detection.

Such combination of the software and hardware redundancy has been proven to successfully detect a variety of attacks on the sensor, controller, vehicle dynamics, actuator, and controller operating systems [4]. The attack detection must combine control algorithms such as state-estimation, fault detection and diagnosis, fault tolerant control parameter and controller estimations to detect CPS dynamic changes. Specifically, it detects changes in the original system by comparing instantaneous outputs in real-time, while it is shielded from attackers. The entire diagnosis process can cover both the cyber and physical domains. A smooth variable structure filter (SVSF) was used to estimate system states and identify system parameters, which is proven to be robust to model uncertainty and noise. The system diagram of such an approach is shown in Figure 1.4.

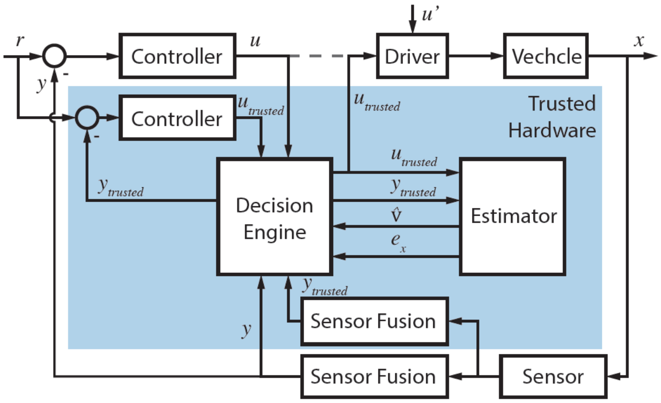


Fig. 1.4 System diagram of the FDI approach using hardware redundancy (reprinted from [4])

Raw sensor data are extracted and used to compare with the feedback from the original system to determine if the sensor fusion result or code in the original system has been modified. The decision engine will integrate the error between the two measurements within a fixed time window and identify sensor attacks using thresholds. The sensor fusion results are also fed to its internal controller and SVSF-based estimator for further security diagnosis and attacks detection.

The sensory system is also critical for CPS safety. Recent advances in adversarial studies demonstrated successful sensing fault generation by targeting the physical vulnerabilities of the sensors. A complete CPS sensor safety design must contain both an fault detection and isolation (FDI) unit and a fault recovery (FR) function to tolerate the detected flaws. When faults are identified and isolated by the FDI, the recovery logic should then be able to maintain the correct state with as much stability as possible using the remaining incomplete sensory systems.

As CPS sensors, the actuation system is also vulnerable and can be easily compromised via similar cyber and physical domain strategies. Actuator failures not only affect the normal operations, due to the implicit dynamics from the actual system, they also introduce the need of FDI design to distinguish the exact sensing and actuation faults. A flawed sensor may induce multiple state anomalies simultaneously and CPS may yield a similar abnormal action to two completely different types of failures (e.g., sensor or actuator failure which complicates pinpointing the failure source). Hence, to achieve proper FDI capability, the inherent coupling of the CPS dynamics must be considered. Further, when multiple CPS states malfunction simultaneously, it is hard to identify the exact failure sensor. The cascading effect may also have to be considered. For example, the high-level sensor abnormalities can affect low-level sensing in cascade for UAVs (e.g., attitude twitching can be subject to the frequent loss of position feedback).

One major circumstance to help sensor recovery is the fact that in most cases, excluding the most catastrophic, all the onboard sensors cannot be rendered defective at the same time. It is hard to compromise multiple sensors simultaneously because they typically measure different physical terms and possess distinct working principles, communication methods, and signal bandwidth.

To recover the sensor readings, installation of a backup sensory system is the most widely used approach. Through a simple comparison and replacement, this hardware redundancy is effective against the traditional software-based sensor faults and attacks, such as numeric error, trojans and data spoofing. However, this approach is not sufficient when the CPS encounters some well-designed attacks that concern both cyber and physical properties of the targeted sensors. This is because the redundant sensors exhibit the same physical vulnerabilities as the original ones. For example, a redundant attitude sensor would be incapable of nullifying the effects of resonating the inertial sensors via external excitation. It is highly likely both the original and redundant sensors would fail.

As an alternative to a redundant hardware approach, the redundancy-free methods for CPS sensor FDI and fault recovery (FR) as reported in [11] make the most sense. Figure 1.5 shows the system diagram, which consists of a fine-grained sensor FDI architecture and a sensor complementary FR in parallel. For fine-grained FDI, a basic

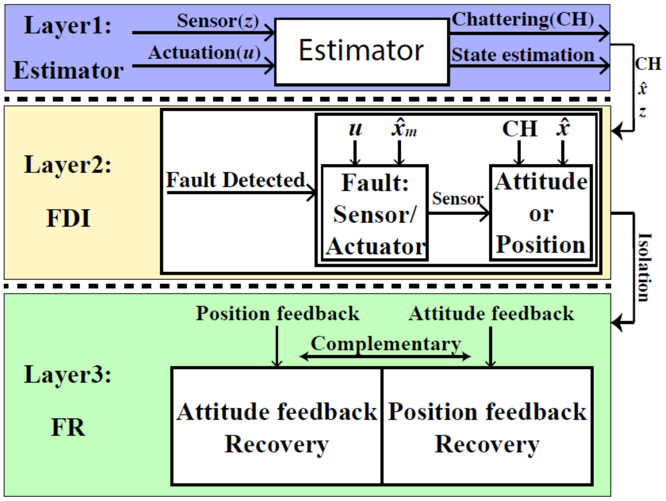


Fig. 1.5 FDI/FR without hardware redundancy (reprinted from [11])

state estimator for a rough early warning of faults combined with the un-measurable actuator state and modeling uncertainties are utilized. Instead of adding auxiliary sensors, the method uses the original sensor arrays and leverages complementary sensor estimations for FR implementation.

The FDI design is based on smooth variable structure filter (SVSF), which is a sliding-mode-based state estimator with a prediction-correction workflow. In one iteration, a model-based prediction function generates *a priori* state estimation first, then a discrete corrective action is taken by adding a corrective gain. The corrective gain is not only used to guarantee the stability of the estimator but also rectify the bounded estimation error robustly. Subsequently, the updated *posteriori* estimation and state measurement carry out the next iteration. The fault is detected by examining sensor FDI procedure-residual check (i.e., the discrepancy between estimation and actual sensor reading). If the residual rises beyond a certain threshold, a fault is supposed to be present and will be reported to begin the recovery.

The sensor fault recovery without hardware redundancy is CPS domain specific. For example, for UAVs, due to the geometric correlation of the vehicle dynamics, position and attitude feedback can be used to compensate each other. During the recovery process, with the fine-grained FDI, the compromised sensor reading is rejected and compensation from other trusted sensors will be utilized. For example, in case of an inertia sensor failure, position information can be used to derive an alternative attitude for flight control. When the UAV loses its position feedback, the inertia measurement can be utilized to compensate position drift.

In summary, reference models provide feasible and robust means for state estimation, behavior prediction, discrepancy checking, decoupling of sensor and actuator faults, and diagnosing multiple faults and accurately isolating the source of faulty elements, thus offering a well-grounded base for building security and resilience in cyber physical systems.

1.2.2.5 Defense: Vulnerability Prevention

To prevent against the first stage of a cyber exploit (see Section 1.2.2.2), CPS software needs to be devoid of any exploitable vulnerability. Fault or vulnerability avoidance generally falls within the first quadrant (P(S), C(S)) of Figure 1.2. CPS software can be analyzed against exploitable vulnerability, and the location where a vulnerability is identified will be *hardened* with security checks or assertions. Software vulnerability analysis generally uses both static (e.g., symbolic execution) and dynamic analysis (e.g., fuzzing) tools for finding exploitable vulnerabilities. *Formal verification* is another approach for assuring that the software is devoid of flaws or vulnerabilities. We will describe hardening and formal methods in what follows.

Security hardening

Vulnerability analysis and hardening is usually performed in several steps: (i) static software program analysis, (ii) instrumentation, (iii) symbolic execution, and (iv) fuzzing with dynamic tracing/feedback-guided fuzzing with sanitizing.

Software program analysis includes combination of static analysis (e.g., dependency analysis, program slicing, etc.) and a symbolic exploration of the program's state space (e.g., "can we execute it until we find an overflow?" or "let's execute only program slices that lead to a memory write to find an overflow."). Symbolic execution also takes advantage of instrumentation for precise results.

Fuzzing is a form of software testing where an application is run with random (potentially malformed) inputs while monitoring the runtime for unexpected behaviors, e.g., crashes, memory exhaustion, or infinite loops. There are generally two types of fuzzing: (i) *Blackbox fuzzing* (i.e., fuzzing with no knowledge about the target application) may not be effective in many cases as most inputs are likely to explore very shallow code paths. This severely limits the fuzzing ability to uncover bugs in deep parts of the code. (ii) *Coverage-guided fuzzing* tackles this problem by using program traces generated by the inputs as a feedback mechanism to tailor future inputs to the fuzzing target.

In essence, fuzzing depends on program crashes to detect and report bugs. Consequently, bugs that do not trigger crashes are not caught through fuzzing. Therefore, for effective fuzzing, software must be instrumented with sanitizers (e.g., a memory checking code such as memory leaks and initialization, heap and stack overflows, illegal accesses, etc.) either at compile time or at the binary level.

Program analysis and instrumentation can be performed for newly developed software during CPS software code compilation, or, on the available binary code for legacy software. For the former case (i.e., compiler-time analysis and hardening), the mainstream software build tools (e.g., GCC and LLVM) provide extensive interface and framework for analysis and code optimization in its Intermediate Representation (IR) form during code compilation. For the latter case (i.e., binary analysis and rewriting), analysis in the form of symbolic execution with various static analyses on binaries is performed as three distinct steps: (i) loading a binary into the analysis program, (ii) translating a binary into an IR, and (iii) performing the actual analysis.

The rewriting part for instrumentation and hardening presents a few difficult challenges. Specifically, dis-ambiguating reference and scalar constants, so that a program can be "re-flowed" (i.e., having its code and data pointers adjusted according to the inserted instrumentation and data section changes) is a major challenge. During assembly, labels are translated into relative offsets or relocation entries. A static binary rewriter must recover all these offsets correctly. There are three fundamental techniques to rewrite binaries: (i) lifting the code to an intermediate representation, (ii) trampolines, which rely on indirection to insert new code segments without changing the size of basic blocks, and (iii) reassemblable assembly, which creates an assembly file equivalent to what a compiler would emit (i.e., with relocation symbols for the linker to resolve).

Lifting code to IR for recompilation requires correctly recovering type information from binaries, which remains an open problem. Trampolines may significantly increase code size and do not scale very well. Consequently, we believe that *reassemblable assembly* is the most promising approach, which creates assembly files that appear to be compiler-generated (i.e., do not contain hard-coded values but assembly labels). Symbolizing the assembly allows security-oriented rewriters to directly modify binaries, which is similar to editing compiler-generated assembly files. Once modified, the symbolized assembly files can be assembled using any off-the-shelf assembler to generate an instrumented binary.

There are a number of security mechanisms and sanitizers to harden programs written in unsafe languages, each of which mitigating a specific type of memory error. It includes various memory checking techniques, undefined behavior monitors, control flow integrity trackers, temporal safety enforcers, etc. The major problem is that the execution slowdown caused by various security mechanisms is often non-linearly accumulated, making the combined protection prohibitively expensive.

One of the most viable approaches to mitigate this problem is to use *security diversification* consisting of N variants that are both functionally identical in normal situations and behaviorally different when under attacks. Hence, although each program version may be vulnerable to certain types of attacks, the security of the whole system relies on the notion that an attacker has to simultaneously succeed in attacking all variants in order to compromise the whole system. In addition, different and even conflicting security mechanisms can be combined to secure a program while reducing the execution slowdown by automatically distributing runtime security checks in multiple program variants [12]. This can be achieved by making sure that conflicts between security checks are inherently eliminated and execution slowdown is minimized with parallel execution. The N-version execution engine synchronizes these variants so that all distributed security checks work together to guarantee the security of a target program. The notional workflow diagram is shown in Figure 1.6.

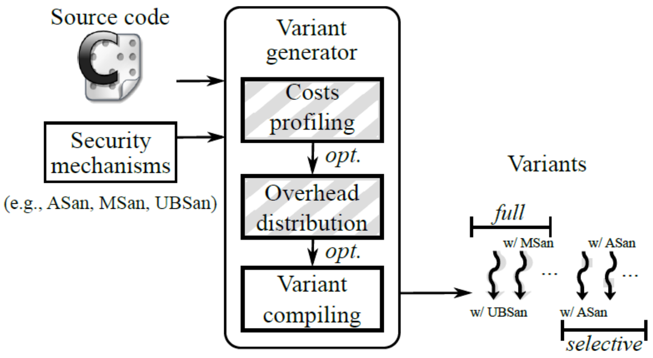


Fig. 1.6 Variant Generation Workflow (reprint from [12])

Formal Methods

While program analysis tools are indispensable to find vulnerabilities, they essentially explore part of the state and execution space, and hence can never provide complete guarantees. Formal methods provide complementary means for vulnerability prevention. Generally speaking, formal methods are system design and analysis techniques that use rigorously specified mathematical models to build software and hardware systems. In other words, formal methods use mathematical proofs as a complement to system testing (e.g., fuzzing) in order to ensure correct behaviors.

Using the terms mentioned in Section 1.2.1, consider a binary executable as the *thing*. In general, the *Concrete Model* or M_c is comprised of (but not limited to) an instrumented C or C-like program (source). The concrete model represents the actual executable but is more generic, meaning some properties are present in the model that are not reflected in the binary. An *Abstract Model* or M_A (for reasoning) is comprised of formal constructs such as Hoare's logic (or its variants such as separation logic, higher order logic, etc.). It can be derived from the binary or M_c and captures all properties in binary or M_c but is more generic and might include properties not in binary or M_c. In an ideal world, no error in lifting or abstraction is made and the three concepts (executable, M_c, and M_A) are all identical. In reality, however, this is not the case and much research has been conducted to shorten the gaps between them. One notable methodology is counter-example guided abstract refinement (CEGAR) where M_A is iteratively checked against a given property and refined if the check fails [13]. Most of the formal verification efforts adopt a similar model refinement approach.

Over the last decade, the understanding of formal methods and development of tools have improved to the point where formal verification of real-world software has started to become feasible. Examples include functional correctness proofs of microkernels and cryptography libraries. Formal methods have also been used to identify deep vulnerabilities in software, revitalizing the field of program analysis. With respect to vulnerability prevention, seL4 is the first formally verified microkernel with a functional correctness proof of the abstracted source code against the specification, effectively asserting the absence of typical programming errors such as null pointer dereferences, buffer overflows, and arithmetic exceptions [14]. The development of seL4 was supported by the DARPA High-Assurance Cyber Military Systems (HACMS) program, which aimed to create technology for the construction of high-assurance cyber-physical systems, where high assurance is defined to mean functionally correct and satisfying appropriate safety and security properties. Since it's original proof over a decade ago, seL4 has seen reasonable successes in both continual development and early adoption. For example, due to the offered higher-level of confidence for assurance, seL4 was selected by the AFRL Agile and Resilient Embedded Systems (ARES) program to serve as the separation kernel providing memory allocation and isolation based on the hardware memory management support.

However, while formal methods provide a rigorous way for vulnerability prevention, it is important to point out that the proofs are usually carried out between *models* (e.g., abstract and concrete models), as opposed to against the *thing* or binary executable in this case. For example, just because seL4 is formally verified does not necessarily mean the binary executable runs exactly as expected in the specification on the target computer architecture. Additional binary level assertion is still needed if the purpose is to provide execution assurance directly on the computer hardware. Again, this is a reminding and cautionary tale related to the *thing* and its *models*, as discussed in Section 1.2.1.

1.2.2.6 Defense: Vulnerability Tolerance

Most safety-critical systems utilize some type of redundant architecture to deal with faults. Examples include hot backups; dual, triple, or quad-redundant architecture; or Byzantine fault tolerance where assumptions about the fault conditions are random, and faulty replicas may behave arbitrarily. Fault tolerance provides a means to automatically deal with faults and recover from them. Cyber attacks, however, will drive fault tolerant system into *common mode failures* (see Section 1.2.2.2). The challenge is how to retrofit existing fault tolerance architecture to rectify faults caused by cyber attacks.

A typical cyber physical system offers certain properties and advantages one would not find in a general IT system. This is because the physical aspect allows for a certain degree of predictability in the behaviors of the system.

- **Periodicity:** The cyber subsystem that directly interacts with the physical plant runs in continuous cycles. For example, throughout its execution the controller reads values from sensors, calculates the error correction signal, and writes out actuator values. For the commonly used industrial controllers, Programmable Logic Controllers (PLCs), this is called the *scan cycle*.
- **Inertia:** Any physical subsystem of a CPS must obey the laws of physics and physical systems inherently have inertia. The scan cycle of a controller is typically engineered to be fast enough such that an issue in a small number of cycles will

be dampened out by the existing inertia. The cycle frequency is set depending on the system but common values vary anywhere between 1 Hz and 1 kHz.

Due to this predictability offered by inertia and periodicity, anomaly detection approaches can be naturally used to detect anomalies and threats in the system. A resilience strategy can also be developed to detect attacks by monitoring actions such as subverting control flow, reprogramming controllers, or overriding sensors that are out of the normal operation ranges.

Cyber vulnerability (and attack) tolerance does not rely on the need for software to be devoid of vulnerabilities. Instead, it assumes that unknown vulnerability exists within the software and strives to maintain the safety and normalcy of system operation regardless. Vulnerability tolerance methods focus on the second stage of cyber exploits (see section 1.2.2.2) and will generally have to perform timely recovery within the limited time afforded by the physical systems' inertia, as the first stage of cyber exploits may have already occurred and the cyber systems may have been corrupted. In what follows we will describe some example techniques, tools and frameworks for providing vulnerability and attack tolerance, empowered by inertia and predictability unique in cyber physical systems. These include Software Brittleness, Byzantine Fault Tolerance++ (BFT++), You Only Live Once (YOLO), and CPS Cyber Resilience Architecture (CRA).

Software Brittleness

For some certain types of critical cyber physical systems, avoiding operating in degraded or compromised state is of paramount importance, and fast program exit and re-start (called *software brittleness*) is required when a cyber attack succeeds and the program control is lost. Examples include Industrial Automation and Control Systems(IACS), Supervisory Control and Data Acquisition (SCADA) control systems and devices, Programmable automation controllers (PAC), remote terminal units (RTU), Master terminal units (MTU), intelligent electronic devices (IED), etc. Software brittleness is a novel concept enabled by the new design space, i.e., the fourth quadrant with P(S) and C(U) as shown in Figure 1.2b. Essentially, the inherent physical inertia allows enough room for cyber components to reconstitute themselves via fast crash-and-recovery.

Code randomization/diversification for software brittleness can be implemented at either pre-distribution or post-distribution stages. Both types of diversification (i) provide the level of code diversification sufficient to guarantee that an attack that succeeds in the original program will fail in the variants, and (ii) assure prompt attack discovery through self-monitoring capabilities of the diversified code. Using N-voting system with simultaneously running multiple generated variants will assure prompt discovery and recovery. There is an integrated set of diversification techniques available at both the source and binary code levels against most known attacks (e.g., memory corruption, code injection and re-use, control flow hijacking, information leaks, etc.). The conceptual approach toward software brittleness, called Binary code Randomization for Attack Sensitive Software (BRASS), is shown in Figure

1.7. It has been demonstrated that this approach provides prompt attack discovery and program abort & recovery with low performance and size overhead [15]. In the CPS context, software brittleness can be included in controllers, for example, and managed through some vulnerability/fault tolerance framework, which comes next.

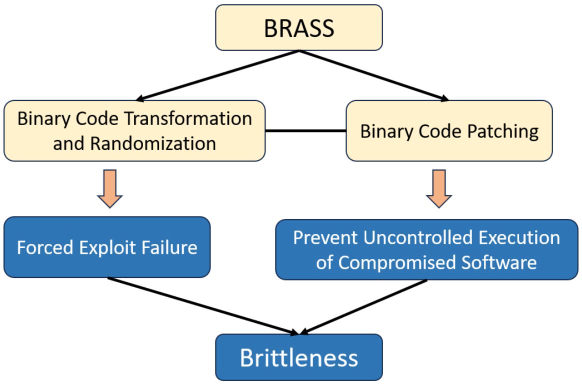

BRASS
Binary Code Transformation and Randomization
Binary Code Patching
Forced Exploit Failure
Prevent Uncontrolled Execution of Compromised Software
Brittleness

Fig. 1.7 The conceptual approach toward software brittleness

Byzantine Fault Tolerance++ (BFT++)

BFT++ is a family of cyber attack resilience methods that rely on the periodicity of CPS and the physical inertia to tolerate cyber attacks [16]. The initial concept of BFT++ was developed by the Office of Naval Research (ONR). It operates in the fourth quadrant (P(S), C(U)) in Figure 1.2b. MITRE Corp. maintains a reference design for BFT++ for the NAVY and DoD in general. BFT++ has been demonstrated to withstand US-NAVY sponsored ”Hack The Machine” hackathon.

Figure 1.8 illustrates the main components of the BFT++ design. It is built over the classical BFT systems. *Artificial diversity* in the form of diversified software compilation or diversified processors (ISAs) is used to break common-mode failure (C0 & C1), and delayed input sharing (delay queue) is inserted into the input of the backup unit/controller for stateful (warm) recovery. For a CPS system that does not demand stateful recovery, the delay queue and backup unit (C2) can be omitted.

BFT++ uses artificial software diversification applied to existing code. As an alternative, two (or more) distinct processors of different instruction set architectures can also be used to provide diversity. When used in combination with traditional fault tolerance architecture, this is also effective at absorbing (and tolerating) cyber faults. Note that attackers only have one opportunity per scan cycle to provide corrupting inputs. Diverse replicas will have different code layouts, making it almost impossible for attackers to inject malicious code that works across all replicas simultaneously. Due to the real-time nature of the periodic control loops, synchronization across replicas is built-in, and an attack can be detected if the program results vary across replicas or if timely responses are not provided.

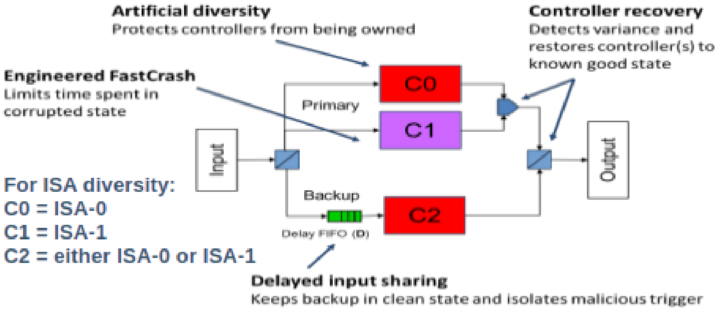

Artificial diversity
Protects controllers from being owned
Controller recovery
Detects variance and restores controller(s) to known good state
Engineered FastCrash
Limits time spent in corrupted state
C0
C1
C2
Primary
Backup
Input
Output
Delay FIFO (D)
For ISA diversity:
C0 = ISA-0
C1 = ISA-1
C2 = either ISA-0 or ISA-1
Delayed input sharing
Keeps backup in clean state and isolates malicious trigger

Fig. 1.8 General Design of BFT++ Cyber Attack Resilience Methods for CPS

The next crucial step is enabling the system to recover. Diversity within the system can make it more fragile, so fast-acting and automated recovery must be employed to counterbalance this. Without recovery, the attacker could maintain control of one (compromised) replica and leave the others crashed – a clearly unacceptable state.

Crash detection is the first part of the recovery process. Ideally, we want to detect a potential compromise via a crash of one of the diversified replicas as soon as possible. In our case, a replica failing to produce timely output by the end of the epoch is considered to have crashed. This serves as a canary that there has been a compromise.

Next, a small message queue is employed in front of one of the replicas (henceforth referred to as the ”protected replica”). This is key, because when a potential compromise is detected (via the crash detection), the message(s) triggering said crash are trapped in the queue before reaching the protected replica. Upon crash detection, this queue can be flushed removing the offending messages. While this introduces a small delay to the protected controller, the physical inertia of the system allows BFT++ to absorb this without impact to the real-time operation.

Finally, recovery begins, and the state of the replicas are restored. Restoring from a checkpoint is possible but requires much resources to handle the overhead of saving checkpoints as well as a way to deal with the staleness of state upon a restore. Instead, the strategy advocates designating one or more replicas as backups and time-delaying them, so they process inputs one or more cycles behind the primaries.

This method for cyber resilience has allowed older control systems to identify cyber attacks during their normal operation, automatically triggering a quick and efficient recovery process. However, there is still a concern that attackers may exploit this system behavior to launch an availability attack. While we have prevented any exploit from affecting the system’s integrity, it is possible for a known vulnerability or bug to trigger the recovery process and cause the system’s availability to be compromised. To address this issue, we designed a mitigation strategy known as ”Shims” that filters out any malicious inputs that cause the recovery architecture to send a crash signal. By implementing shims at the input point for the controllers, replaying an exploit after it has already been used against the system will be prevented.

For a particular BFT++ implementation [16], the architecture has three redundant diversified controllers operating in a traditional fault-tolerant architecture. The artificial diversity makes it difficult for a cyber attacker to compromise all controllers with the same malicious input. Although an exploit may be successful against one replica, it will cause the diversified replicas to crash. Next, it incorporates delayed input sharing (e.g., FIFO message queue) to trap bad messages before reaching a ”protected controller”. This introduces a delay to the protected controller, but ensures the system to continue operation and to be reconstituted after the cyber exploit. The recovery timing of the system is governed by several timing parameters, such as Tcrash, Tsc, D, Td, and Tr. Tsc, Td, and Tr are system parameters, and D needs to be appropriately set for automated recovery to be possible. The two critical points that determine the system’s recovery timing are the *brittleness* of the controllers and how quickly the system can restore a controller to the normal state. The physical subsystems with higher inertia are generally more tolerant of losing control signals for a short time. In general, the following relationship between these parameters must hold for BFT++ to be applicable to a legacy cyber physical system [16]:

$$Tcrash \leq D * Tsc \leq Td - Tr$$

Parameters	**Definitions**
Tcrash	Time between attack and crash
Tsc	The scan cycle period
D	FIFO queue length (number of slots)
Td	Maximum control loss tolerable by physical systems
Tr	Recovery latency

You Only Live Once (YOLO)

YOLO is another CPS cyber resilience method that relies on the physical inertia to withstand cyber attacks. YOLO and its variants use periodic restart and does not require any redundant controller [17]. It also operates in the fourth quadrant (P(S), C(U)) in Figure 1.2b. YOLO was developed at Columbia University under the sponsorship of the Office of Naval Research.

Figure 1.9 depicts the design philosophy for YOLO. YOLO implements periodic restart to limit the duration of potential compromise. An adversary who managed to compromise the system only has control over the system for a maximum duration of the restart period. The YOLO restart period can be designed to be short enough to prevent an adversary to achieve persistence within the CPS system, while still within the tolerable region provided by the physical systems’ inertia. During each restart, the controller is reset to its ‘clean’ state by loading its software from a read-only module and clearing out all the volatile memory. YOLO also implements software diversity

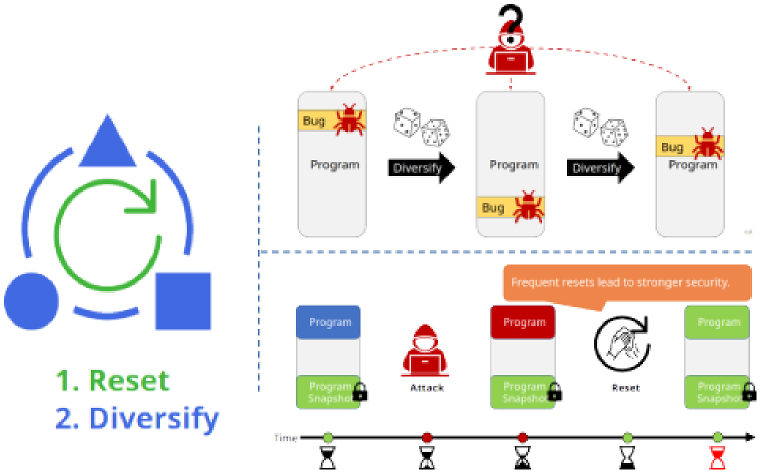


Fig. 1.9 General Design of YOLO Cyber Attack Resilience Methods for CPS [17]

after each restart to ensure that the attacker cannot exploit the same vulnerabilities. YOLO has been demonstrated to be practical for automotive engine management unit, drone controller and a missile launcher.

In YOLO, the restating latency and state recovery time need to be within the range that the inertia of the physical systems can tolerate. Proper engineering analysis and design is required to accelerate the restarting process and to avoid lengthy reboot and initialization latency of the cyber system. YOLO does not require replication, and hence it is cheaper to implement than BFT++. However, its protection is not as deterministic as that of BFT++, and current version of YOLO does not support stateful (warm) recovery.

CPS Cyber Resilience Architecture (CRA)

Existing CPS cyber resilience architectures, including BFT++ and YOLO, have been analyzed and summarized into a timing-based formulation framework [18]. Within this framework, safety analysis and computation of control policies and design parameters can be performed for each pair of CRA method and CPS application.

The framework relies on the insight that the cyber subsystem operates in one of a finite number of modes. It defines a hybrid system model that captures a CPS adopting any of these architectures (CRAs). Analysis within this framework uses the transition model of the hybrid system to derive architecture-agnostic sufficient conditions for control policy and timing parameters that ensure safety of the CPS. The analysis will then formulate the problem of joint computation of control policies and associated timing parameters for the CPS to satisfy a given safety constraint and derive sufficient conditions for the solution. Utilizing the derived conditions, they provide an algorithm to compute control policies and timing parameters relevant to the employed architecture. The framework efficacy has been demonstrated in a case study involving automotive adaptive cruise control. The study was performed for each of the CRA methods in their framework, and proved that the algorithm converges to a feasible solution under certain conditions.

Figure 1.10 visualizes the operation of a drone employing YOLO, under continuous cyber attacks. It shows three operation zones: desired operating zone, zone of tolerance, and danger zone. The drone is expected to operate in the desired safe zone, and danger zone can only be safely entered when the vehicle is in normal mode, otherwise catastrophic crash may occur. The vehicle (drone) can be safely restarted and recovered within the tolerated zone. Safety cannot be guaranteed if the drone enters tolerated zone in a corrupted state.

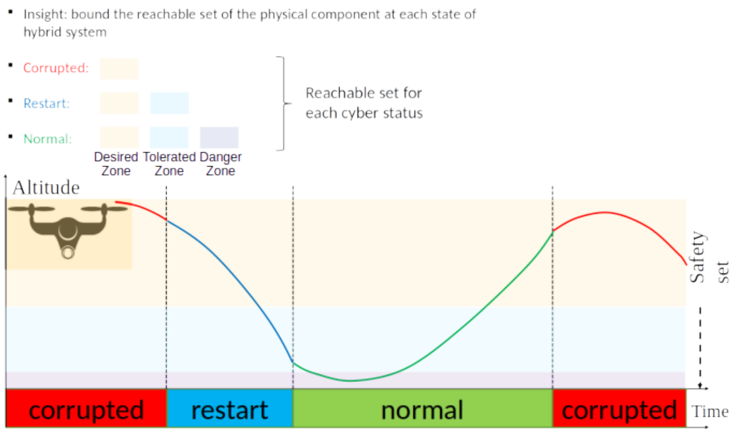

Insight: bound the reachable set of the physical component at each state of hybrid system
Corrupted:
Restart:
Normal:
Reachable set for each cyber status
Desired Zone
Tolerated Zone
Danger Zone
Altitude
Safety set
corrupted
restart
normal
corrupted
Time

Fig. 1.10 An illustration of CPS Safety Analysis within the Framework [18]

1.3 Machine Learning and CPS

In this chapter, the terms *machine learning* and *neural networks* will be used interchangeably. At the most general level of interpretation, machine learning is a super-set of statistical machine learning (i.e., neural networks). The term *machine learning* in general also includes learning heuristics and other logical and formal form of learning mechanisms. Contemporary use of the term *machine learning* generally refers to various forms of neural networks, which are often considered as surrogates for formal/logical control algorithms in the CPS context.

Robotic devices and vehicles can perform some of its functions using machine learning and especially reinforcement learning (RL), e.g., RL for drone fault recovery [19]. Training neural networks may be performed by guiding the robotic devices to function within its physical environment. Training may also be conducted in the virtual simulation environment [19], where the sensory and control input as well as the actuators and their dynamics are simulated using physics models of the actuators, sensors and the physical environment. The use of physics models for training a CPS system is generally safe as the laws of physics are universal, relatively complete, consistent and context insensitive.

Machine learning may also be used for correlating various monitored and logged events in cyber physical systems. It helps correlate cyber events such as network events, activation of computing events, sensed and computed parameters' values, etc., with observed physical and environmental events. Trained this way, machine learning models the operational behavior of the cyber physical systems and can be used to highlight unexpected behavior and anomalies. The use of machine learning in this case is inherently incomplete. While well trained machine learning algorithm/model is expected to generalize and cover the CPS operation space, there is no practical, assured way to ascertain that it covers all of the possible cases of the application/CPS-operation, e.g., corner and unexpected cases. Such machine learning model will produce false positives and false negatives. The quality of machine learning output (prediction) is significantly dependent on the methods, the quality of data, and models used for training. However, with proper operator due diligence and supervision, the deployment of machine learning can significantly improve the safety and security of CPS operation, as a complementary means to the traditional model-based mechanisms.

1.3.1 Enhancing CPS robustness with Machine Learning

It has been well established that the traditional cyber techniques in software and firmware can no longer sufficiently protect the system and ensure safe operation of the cyber physical systems when attacks are launched against the physical components of the CPS, such as signal spoofing or using sound wave to resonate the IMU sensors. As a result, undesirable performance or even loss of control would occur. Given that attacks/faults cannot be fully prevented, fault/vulnerability tolerance and CPS resilience and recovery strategies are required.

Traditionally, there are two types of fault-tolerant control: passive fault-tolerant control (PFTC) and active fault-tolerant control (AFTC). AFTC has a fault detection and diagnostics (FDD) component to identify the source of the fault, reconfigure a controller, and compensate for such fault. The FDD component is usually an observer and can generate residual signals to indicate the fault. Both sensor and actuator attacks or failures can be detected with system models. Meanwhile, PFTC does not have an FDD mechanism, but aims to improve the controller's robustness and tolerate the fault condition or attack. AFTC can pinpoint the fault and act accordingly, but if the FDD is not designed with care, the implementation could lead to delay in detection or false positives and greatly affect the performance. While PFTC cannot isolate faults, they could potentially achieve more robust performance.

Machine learning (ML) and reinforcement learning (RL) have been explored in developing FTC strategies. However, most of the ML/RL methods were only evaluated in simulation, and their real-world performance is unknown. Deploying reinforcement learning policies onto real systems in this case is extremely challenging since training has to be performed in simulation before trained models being transferred to real cyber physical systems to recover from sensor and actuator faults.

Reference [19] demonstrates that RL-based policy trained in simulation can indeed be transferred to real unmanned aerial vehicles to recover from sensor and actuator faults. Unlike traditional FTC, this policy does not require fault detection and diagnosis (FDD) nor tailoring the controller for specific attack scenarios. Instead, the policy runs simultaneously alongside the controller without the need for fault detection and activation. When the CPS is operating normally, the policy generates no or minimum control command adjustment and does not interfere with the operation. When the fault condition arises, or the CPS is under attack, the policy takes the state inputs and generates appropriate actuator command adjustment with little or no delay to compensate for the fault/attack condition.

For simulation, identical sensor fusion and control algorithms must be implemented so the closed-loop dynamics of the vehicle in the simulation can approximate the real vehicle. The approach allows for simulation validation statically through open-loop and closed-loop tests. The fault resilient policy optimization is formulated with a standard reinforcement learning problem, where the agent is the quadcopter and the environment is the simulated world. The attacks were implemented by replacing the actuator signal or the sensor value with a random number.

The training takes place and policy is implemented during dynamic simulation on existing legacy system through minimally intrusive software retrofitting. The control algorithm is used in a dynamic simulation during which a fault-tolerant policy is optimized using reinforcement learning to maintain CPS control under various simulated attacks. The RL-assisted fault-tolerant control workflow is shown in Figure 1.11.

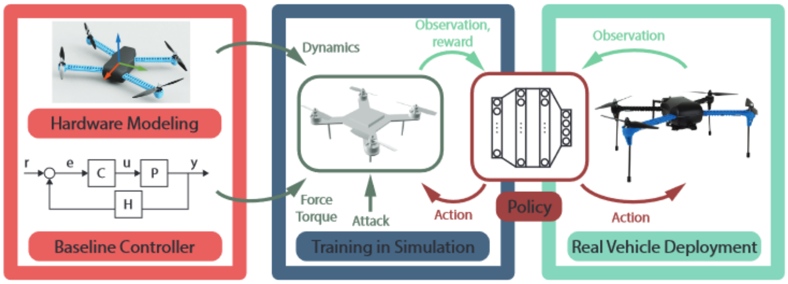


Fig. 1.11 RL-assisted fault-tolerant control workflow, derived from [19]

Since the system dynamics is largely deterministic, reference [19] uses the actor-critic deep deterministic policy gradient (DDPG) algorithm for training. Fully connected multi-layer perceptions (MLPs) serve as the function approximator for policy representation. MLP splits between a nonlinear control module and a linear control module. Intuitively, nonlinear control performs forward-looking and global control, while linear control stabilizes the local dynamics around the residuals of global control. This improves training sample efficiency, final episodic reward, and generalization of learned policy, while requiring a smaller network and being generally applicable to different training methods. Only nonlinear fault-tolerant policy needs to be learned. This approach views the policy as an optimized FDD and FTC control approximated by a neural network. It can be leveraged (with the proper simulation setup) for a variety of cyber physical systems with a learned policy designed as an add-on to other closed-loop mechanisms.

Another use of ML is to function as a surrogate ("Digital Twins") and to be used as a reference to detect anomalies and cyber disruption. In this role, the ML model will correlate cyber and physical events and flag any inconsistencies, plausible faults or anomalies. Such approaches extend machine learning (ML) methods for analyzing system logs of CPS and identifying the key CPS entities to reconstruct the critical steps of a plausible attack. Forensic analysts collect diverse system logs from multiple CPS components. The massive volumes of logs are often analyzed offline or monitored in real-time to debug system failures and identify sophisticated threats and vulnerabilities. There are several techniques being developed to extract features/sequences from logs to automate intrusion and failure detection and to discover associations among disparate log events through event correlation.

Working with the text logs requires integration of natural language processing (NLP) and deep learning techniques into data provenance analysis to identify attack and non-attack sequences. The typical steps include (i) processing system logs; (ii) building optimized causal dependency graph, from which the semantically augmented event sequences are constructed; and (iii) learning a sequence-based model that represents the attack semantics to recover key attack entities describing the attack story at inference time. The key challenges for such solutions are (i) additional overhead on a running system, (ii) integration of diverse logs, (iii) scalability of the large and complex causal graphs, (iv) accuracy of constructed sequences models, and (v) efficient automation.

One of the promising approaches is based on the assumption that the crucial steps of different attacks in a causal dependency graph may share similar patterns. This allows for identification of key attack steps through an attack symptom event, based on those sequences that share semantically similar attack patterns to the ones it had pre-learned. Such knowledge helps to substantially save time when investigating large causal graphs and helps in constructing the attack story from a limited number of attack symptoms.

1.3.2 Roles and Pitfalls of AI in CPS

As the complexity of automation increases, the roles machine learning may play in CPS are also expected to grow. The use of machine learning often requires an extensive set of labelled data for training, and the curation of this large, labelled data set is often problematic. While data can be scraped from the Internet, labelling them requires tedious manual effort, and is often outsourced to third-world country or Mechanical Turk (Amazon). It can be a very expensive proposition.

Fortunately, many CPS operations are governed by physics, with formulas and models that have been developed and proven over decades. Due to the computation limitation of many CPS devices, and the potential complexity of the interaction among physical phenomena, it is often not practical to deploy detailed physics models as reference to the operation of the CPS infrastructure. A surrogate – a (computationally) lighter weight machine learning algorithm/model – can be trained with labelled data generated by these complex physics models and practically deployed as the reference model. This surrogate model trades off precision and determinism/correctness with computation cost.

However, machine learning for cyber components at the level of software execution is quite challenging. Unlike natural language, image and video processing, there is no public, large-scale, comprehensive, and well-labelled data set that researchers can use for evaluating the efficacy of machine learning for cyber security and resilience. Research works in this area are often forced to develop their own data for training and evaluation. This effort is both expensive and non-comprehensive, limiting the quality and generality of the research effort.

The periodicity and predictability of CPS operation help reduce the overall challenge, as they potentially provide “structures” and “constraints” for the learning problem at hand. In machine learning, *knowledge* about the problem domain and relevant features extracted from the domain knowledge still play a critical role. Properly observing/incorporating physics-based models in the CPS software and machine learning process will help focus the training process, constrain the search space, and enhance the performance of the resulting machine learning model.

Transfer learning offers an appealing way to help reduce the size of training data needed to achieve reasonable performance. Employing transfer learning, one can adopt a suitable pre-trained ML model whose size and structure can accommodate the target problem space. The pre-trained ML model is expected to have its internal weight well configured and distributed, especially for the application it was trained for. It serves as the initial condition and foundation for training the target application. This pre-trained model will then be trained again with labelled data for the new (target) application.

Employing pre-trained model, the required training data is not as large as that of training the machine learning model from scratch. The trade-off is that the configuration and many of the hyper-parameters of the neural networks are not tunable and can incur the computing cost of employing larger than optimum (for the target application) neural networks. Large Language models, e.g., GPT-4, BERT, etc. are powerful examples of pre-trained models for natural language. It is harder to find a suitable pre-trained graph-based machine learning model, as the data encoding for graph neural networks is generally very specific to the application. Fortunately, the training data requirement for various graph neural networks tends to be modest. For cyber components and software execution that are typically represented in graphs, it is still unclear how and how well transfer learning may help in model training with reduced data set.

Generative Adversarial Network (GAN) offers another attractive method in dealing with training data. A GAN consist of two neural networks, the generator and the discriminator. In a GAN setup, the two neural networks contest with each other in the form of a zero-sum game, where one agent's gain is another agent's loss. The generator strives to generate samples that fool the discriminator, and the discriminator strives to accurately detect or classify the generated samples. Both neural networks are trained together and co-evolve against each other. After the initial setup, the generator and discriminator will challenge and train each other in an unsupervised manner. GAN requires minimal if any training data, making it very attractive for domains lacking large-scale, labelled data.

Unwise use of ML, such as employing generative adversarial network (GAN) without properly constraining it with physical models/rules, will likely violate the laws of physics and make it inappropriate or even dangerous to deploy. This is because unconstrained GAN will operate and explore solutions in an (un-grounded) virtual world with a much larger space than that of the physically constrained environment the CPS operates within. As GAN is a very attractive method, it is important to understand the problem space before deploying it. To illustrate, consider two slightly different applications. One is a valid application of GAN, and the other is not.

In the first application, a neural network is being developed for detecting malware (a discriminator). To anticipate 0-day malware, it is trained in a GAN environment – a malware generator neural network is developed and coupled with the malware detector in a GAN configuration, and then let loose (they play against each other). This is an appropriate and efficient way for inoculating the detector (discriminator) against 0-days.

In the second application, a dark-hat is mining for 0-day malware that is guaranteed effective against a target that is defended by VirusTotal. The dark-hat decided to deploy GAN, using a similar set up as the first application above. This is an ineffective solution, and the dark-hat will have false-confidence that his mined 0-day will be effective, for the following reason: his discriminator is not grounded to and does not represent VirusTotal. Developing a discriminator that can become a surrogate to VirusTotal will be very difficult if not impossible. VirusTotal and its evolution is influenced by factors that are not under the dark-hat control. The Dark-hat's discriminator will respond and evolve to the generator challenges in a manner that is independent of VirusTotal, and provide feedback to the generator that does not reflect VirusTotal behaviors. One can speculate that given enough resource and time, one can train a super discriminator that is better at detecting malware than ViriusTotal. However, unless one can prove or have well founded confidence that the superior discriminator is a complete superset of VirusTotal capability (no Malware that VirusTotal can detect the discriminator cannot detect), it still cannot provide the assurance that the synthesized malware will pass detection by VirusTotal.

1.3.3 Future Direction for AI in CPS

As discussed in previous sections, statistical models embodied as neural networks (machine learning) are effective in CPS related automation, including surrogate for control policy, automated fault recovery [19], surrogate as digital twin, anomaly detection, etc.

However, care must be taken in deploying neural networks as they are after all statistical machinery and hence cannot completely capture causality and are prone to make mistakes. Unless the utility property of the application itself is statistical, an error detection and exception handler will be required to detect and mitigate the effect of incorrect neural network results. An application is said to have a *statistical utility* if occasional mistakes are expected and tolerated, as long as their frequency is not too large (below a certain threshold) and the overall performance of the algorithm is still above the acceptable performance level. That is, in an application with statistical utility, only average matters and individual error does not. An application does not have a statistical utility if an individual error/mistakes matters.

CPS is a field where the inertia of the physical systems can tolerate a limited duration of errors. However, an individual CPS is susceptible to prolonged errors. A system of CPS devices provides additional resilience, as long as the effects of prolonged errors in a subset of the system components are generally observable within the systems, and the overall systems adapt to the anomalies, or an operator can be alerted for and rectify the operational anomalies.

Various forms of neural networks and various configurations of systems of neural networks have been deployed in CPS infrastructures. Machine learning will also excel in approximating the modeling and controlling of the behavior of a complex system whose behavior is not easily describable with logic or sets of logic. The role of neural networks and systems of neural networks is expected to grow in CPS and process control & automation in general.

Machine learning excels when the utility of the application itself is statistical, and when the application logic is extremely complex to be completely captured using logic or other formal methods. For this reason, an understanding of the problem's space, property and the environment surrounding the problem is the key for successful application of neural networks and the selection of the particular neural network algorithms. A good understanding of problem space will also help avoid pitfalls described in the previous section.

During our journey of studying the property of a problem or task and their potential solutions, the authors observed that *it is easier to solve problems of statistical nature with statistics*, and vice versa, *it is simpler to solve problems of logical nature using logical process*. This dichotomy is analogous to the dichotomy of frequency domain and time domain in signal processing. There are classes of problems that are simpler to solve in frequency domain, and there are other classes that lend themselves to have natural solutions in time domain. In general however, while less efficient or precise, statistical process can be used to approximate a logical one, and logical process (such as logic in modern digital computer) can emulate/simulate statistics.

A system of machine learning algorithms can be arranged in logical manners or simply feed each other for large scale automation. Properly designed systems of machine learning algorithms may mask or alleviate individual algorithm weakness and provide much more accurate and capable ultimate results. There may also be the case where it is prudent to include logical reasoning algorithms into the systems of neural networks, creating hybrid symbolic and statistical (neural networks) systems.

It can be argued that arranging multiple neural networks in a logical pipeline has already shown the promise of hybrid logical-statistical system design. For example, ONR developed the Learn2Reason concept [20][21], advocating the development of a hybrid system of neural networks algorithms and logic-based reasoning. The development of Learn2Reason was inspired by Daniel Kahneman's system-1 and system-2 concept with respect to thinking fast and slow [22]. Initial description of Learn2Reason [20] suggests a blackboard like implementation where the logical and probabilistic/statistical process interact, however, most of the research mentioned in [21] employed the pipelines structure. Recent news [23] indicates that Google's large language model employs logic, in term of generated program, to solve particular tasks where logical processes clearly surpass statistics, e.g., counting, performing arithmetic calculation, reverse spelling a word, etc. The article also stated that Google was following Kahneman's system-1 and system-2 concept [22] in this work.

A hybrid logical and statistical (neural network) based machine learning is the future. It allows for both statistical process and logical process to do what it can do best, and together they provide a superior performance than that of each individual type (logical or statistical). This hybrid learning system will find its place in CPS and CPS-based critical infrastructure of the future.

References

1. Stouffer K et al (2022) Guide to Operational Technology (OT) Security. *NIST ITL Computer Security Resource Center* https://doi.org/10.6028/NIST.SP.800-82r3.ipd
2. Higgins KJ (2023) OT Network Security Myths Busted in a Pair of Hacks. Available via DARKReading: ICS/OT Security https://www.darkreading.com/ics-ot/ot-network-security-myths-busted-in-a-pair-of-hacks Accessed 6 Sept 2023
3. Kim T et al (2020) From control model to program: investigating robotic aerial vehicle accidents with MAYDAY. In: 29th USENIX Security Symposium., Boston, August 2020
4. Fei F et al (2018) Cross-Layer Retrofitting of UAVs Against Cyber-Physical Attacks. In: IEEE International Conference on Robotics and Automation (ICRA), Brisbane, 21-25 May 2018
5. Son Y et al (2015) Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors. In: 24th USENIX Security Symposium., Washington DC, 12-14 August 2015
6. Learning Introspective Control (LINC). https://www.darpa.mil/program/learning-introspective-control Accessed 6 Sept 2023
7. Faithful Integrated Reverse-Engineering and Exploitation (FIRE) https://defencescienceinstitute.com/funding-opportunity/darpa-fire/ Accessed 6 Sept 2023
8. Lee EA (2016) Fundamental Limits of Cyber-Physical Systems Modeling. ACM Trans. on Cyber-Physical Systems:1–26. https://dl.acm.org/doi/10.1145/2912149
9. Box GEP, Draper NR (1987) Empirical Model-Building and Response Surfaces, Wiley, New York
10. Saltzer JH, Kaashoek MF (2009) Principles of Computer System Design, An Introduction, Morgan Kaufmann, Burlington
11. Tu Z et al (2018) Redundancy-Free UAV Sensor Fault Isolation And Recovery. Preprint at https://arxiv.org/pdf/1812.00063v1.pdf
12. Xu M et al (2017) Compositing Security Mechanisms through Diversification. In: USENIX Annual Technical Conference (ATC'17), Santa Clara, 12-14 June 2017
13. Clarke E et al (2000) Counterexample-guided abstraction refinement. In: International Conference on Computer Aided Verification, Chicago, 15-19 July 2000
14. Klein G et al. (2014) Comprehensive formal verification of an OS microkernel. ACM Trans. on Computer Systems (TOCS) 32.1:1-70.
15. Briskin G, Li JH (2022) Binary code Randomization for Attack Sensitive Software (BRASS), Final Report to the Office of Naval Research, Available to US performers upon request and approval.
16. Mertoguno JS et al (2019) A physics-based strategy for cyber resilience of CPS. In: Proc. SPIE 11009, Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure, Baltimore, 2 May 2019
17. Arroyo MA et al (2019) YOLO: frequently resetting cyber-physical systems for security. In: Proc. SPIE 11009, Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure, Baltimore, 2 May 2019
18. Al Maruf A et al (2023) A Timing-Based Framework for Designing Resilient Cyber-Physical Systems under Safety Constraint. ACM Trans. Cyber-Phys. Syst. 7.3:1-25
19. Fei F et al (2020) Learn-to-Recover: Retrofitting UAVs with Reinforcement Learning-Assisted Flight Control Under Cyber-Physical Attacks. In: Proc. IEEE International Conference on Robotics and Automation (ICRA), Virtual, June 2019
20. Mertoguno JS (2014) Human Decision Making Model for Autonomic Cyber Systems. Journal on Artificial Intelligence Tools. 23.6:1-6 https://doi.org/10.1142/S0218213014600239
21. Mertoguno JS (2019) Toward Autonomy: Symbiotic Formal and Statistical Machine Reasoning. In: Proc. 1st IEEE International Conference on Cognitive Machine Intelligence. Los Angeles, 12-14 December 2019
22. Kahneman D (2013) Thinking Fast and Slow. Farrar, Straus and Giroux, New York
23. Amadeo R (2023) Google's Bard AI can now write and execute code to answer a question., ars TECHNICA (2023) https://arstechnica.com/google/2023/06/googles-bard-ai-can-now-write-and-execute-code-to-answer-a-question/ Accessed 6 Sept 2023

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Jonathan L. Goodall

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

10/26/2024 via FDMS

[redacted]

I am responding to the Request for Information on “a National Plan for Cyber-Physical Systems Resilience” on behalf of the Link Lab, an interdisciplinary research lab focusing on cyber-physical systems with over 50 faculty members and 250 student researchers housed within the University of Virginia Engineering School. I serve as the director of Link Lab and as professor of civil and environmental engineering at the University of Virginia. We support the “whole-of-government research and development (R&D) plan related to cyber-physical resilience across systems that may be local, regional, or national in scope.” Link Lab is devoted to the mission of advancing cyber-physical systems broadly including those related to smart cities, smart health, and robotics where resilience is central. In the context of smart cities and smart civil infrastructure, transformation of these systems from physical to cyber-physical systems presents unique opportunities and challenges, as the February 2024 PCAST Report to the President on a Strategy for Cyber-Physical Resilience captures. Interdisciplinary R&D at universities in partnership with government and industry must be encouraged to advance our Nation’s cyber-physical resilience. New education and workforce development efforts in cyber-physical system resilience are also important, as the PCAST report points out. That is why, with support from the National Science Foundation, we have created a graduate certificate program in Cyber-Physical Systems that graduate students in engineering and computer science can earn alongside their graduate degree in a traditional engineering or computer science discipline. As a member of the Virginia Commonwealth Cyber Initiative, we are working with universities across Virginia to integrate cybersecurity into our research, teaching, and training missions for the Link Lab and the University of Virginia more broadly. We also share the viewpoint in the PCAST report that artificial intelligence (AI) is an important tool for making cyber-physical systems resilient. Some examples of research topics we believe are important toward this goal include the following. AI-Enhanced Predictability: Developing sophisticated AI systems that significantly improve the accuracy and reliability of forecasting, planning, and adaptability in system operations. These advanced AI solutions lead to more robust and efficient operations, reducing uncertainties and increasing resilience to disruptions, ultimately enabling businesses and governments to optimize their scheduling, management, and resource allocation with unprecedented precision. Human-AI Copiloting: Creating intelligent assistive technologies that work alongside human operators, augmenting their decision-making capabilities and productivity. These AI systems act as collaborative partners, providing real-time insights, suggestions, and support to enhance human performance across various domains, from complex data analysis to creative problem-solving. Trustworthy and Energy-Efficient AI: Focusing R&D to advance reliability, energy efficiency, transparency, and ethical integrity of AI systems, ensuring they produce accurate, fair, and safe outcomes while maintaining accountability and fostering confidence among users and stakeholders. This approach aims to create AI technologies that not only perform their intended functions with high precision and minimal environmental impact but also adhere to stringent ethical standards and promote open communication thereby building trust and facilitating responsible adoption across various sectors of society. We support the need for cyber-physical resilience for the vast and growing examples of cyber-physical systems of national, regional, and local significance. We are actively working to advance knowledge and practice in this field at the University of Virginia, while also training future engineers and computer scientists in the unique aspects and considerations of cyber-physical systems. We believe an approach that supplements traditional education in established disciplines with training in the new and emerging field of cyber-physical systems is needed to produce future leaders with the breadth and depth of knowledge needed to advance resilient cyber-physical systems. Jonathan L. Goodall, PhD, PE, F.ASCE Director, UVA Engineering Link Lab Professor, Civil and Environmental Engineering University of Virginia

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Luis Garcia

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to NSF Request for Information on Resilience of Cyber-Physical Systems

Federal Register Notice: 89 FR 78915

Submitted by:
Luis Garcia
Assistant Professor
University of Utah

Date: October 26, 2024

Response to RFI on Cyber-Physical System Resilience

As CPS security researchers, we recognize the pressing need for standardized frameworks to guide resilience efforts across diverse cyber-physical systems (CPS), particularly given the range of resilience needs. Beyond traditional threats posed by malicious actors, CPS resilience encompasses naturally occurring events that stress systems, from extreme weather to unprecedented surges in demand during large-scale events like the Olympics. While these scenarios vary, they underscore the need for resilient CPS to withstand deliberate attacks and naturally occurring system-stressing events. Yet, the industry's approach to resilience modeling remains inconsistent, with a lack of unified standards or widely accessible frameworks to guide this resilience across all infrastructure scales.

Digital Twinning in Industry and the Need for Standardization

Digital twinning has emerged as a promising resilience approach, enabling high-fidelity simulations to project system behavior and address vulnerabilities in real-time. For instance, stadiums have adopted digital twin technology to leverage continuous data insights to anticipate and mitigate operational challenges, from the complexity of event logistics to potential system strain during the Olympics. However, access to digital twin technology is limited for smaller, resource-constrained sectors, where resilience measures are critical yet often unaffordable. This lack of standardization leaves essential systems without the guidance to implement cost-effective digital twins for projecting resilience, exacerbating risks for communities that cannot afford redundancy or specialized technology.

Current industry strategies heavily promote redundancy, as seen in Rockwell Automation's guidelines for Distributed Control Systems (DCS) and the IEC 62443 framework. However, for under-resourced municipalities, redundancy at all levels is not feasible–especially when organizations may not understand the need for resiliency. Instead, resilience models should emphasize sustainable, scalable solutions prioritizing resilience-by-design, ensuring critical systems can withstand high-impact events like natural disasters. Standardizing digital twin technology with adaptable models would help bridge this gap, creating affordable pathways for resilience across various industries.

Labor Shortages and the Future of AI in Safety-Critical Infrastructure

A less visible yet critical resilience challenge is the ongoing labor shortage in safety-critical industries, from utilities to manufacturing. High turnover rates, complex training demands, and difficulty retaining skilled operators—especially with expertise in cybersecurity practices—are compounding vulnerabilities within critical infrastructure. The trend towards using AI solutions as stand-ins for human operators appears inevitable, particularly for maintaining 24/7 operations in environments where retention is difficult. However, AI-driven automation in these settings must be met with robust guardrails, including cybersecurity and operational integrity safeguards, to protect against potential exploitation or failure. In this context, standardized frameworks for integrating AI solutions within CPS resilience models will be key, ensuring that such systems are both adaptive and secure over time.

Cross-Domain Resilience and the Need for Semantic Standards

While frameworks like MITRE ATT&CK for ICS categorize Tactics, Techniques, and Procedures (TTPs), they lack comprehensive support for resilience across domains. Many prior attacks demonstrate that attackers exploit common abstractions and shared physics across critical infrastructure, targeting similar sensors and actuator mechanisms—such as valves, pumps, and dosing processes—used in sectors like water treatment, oil processing, and gas infrastructure. By formalizing these common targets regarding their impact on physical operations across sectors, we can proactively model OT attack semantics, enabling a more fine-grained approach to threat modeling. This approach would equip defenders to anticipate cross-domain vulnerabilities as attackers increasingly focus on operational disruptions that apply universally across CPS domains.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Meera Sridhar, Chenglong Fu, Miao Wang, Robert Cox, Rick Hudson, Ahmed Helmy, Bojan Cukic

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

UNIVERSITY OF NORTH CAROLINA
CHARLOTTE

University of North Carolina Charlotte's response to NSF RFI on Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience

We are writing in response to the National Science Foundation's Request for Information (RFI) on Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience. We appreciate the opportunity to contribute insights and recommendations to inform future research directions and priorities in this critical area. Below, we provide our recommendations for advancing research in this domain.

The University of North Carolina at Charlotte (UNC Charlotte) is a public research university located in Charlotte, North Carolina. Founded in 1946, it has grown into the second-largest institution in the University of North Carolina system, serving over 30,000 students across undergraduate, graduate, and doctoral programs. UNC Charlotte has established itself as a significant contributor to scientific advancement and a driver of regional innovation, preparing students to lead in a rapidly evolving research landscape. UNC Charlotte is known for its emphasis on applied research and innovation, particularly in fields such as cybersecurity, energy systems, advanced manufacturing, and data science. Leveraging its proximity to Charlotte's dynamic financial and technology sectors, the university fosters partnerships with industries to support innovative research and practical applications, The university's connection to Charlotte provides students with unique internship and job opportunities in a vibrant metropolitan area.

The Center for Energy Security And Reliability (CESAR), spearheaded by UNC Charlotte, is a collaborative center (with NC State University and NCA&T University) dedicated to research innovation for the creation of a robust, cybersecure, and carbon-neutral power grid infrastructure through interdisciplinary research, education, and collaborative partnerships. CESAR brings together a multidisciplinary team of researchers, educators, engineers, and industry partners to solve challenging problems and train the future workforce on critical infrastructure cybersecurity and resilience.

Written by: Meera Sridhar, Chenglong Fu, Miao Wang, Robert Cox, Rick Hudson, Ahmed Helmy, Bojan Cukic.

Please note any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the authors' affiliations or the University of North Carolina at Charlotte or the CESAR member institutions.

UNIVERSITY OF NORTH CAROLINA
CHARLOTTE

1. Challenges and Opportunities in Cyber Security and Resiliency for Critical Infrastructure Research

Cyber-Physical Systems (CPS) are naturally complex, with cross-platform, inter-communicating components. CPS such as power grids, water treatment facilities, and transportation control systems, face significant risks from cyber attacks that can lead to catastrophic real-world consequences. The intricate nature of CPS, characterized by numerous interdependent processes and devices, requires extensive manual intervention by highly skilled experts, limiting the feasibility of automated rapid-response mechanisms and potentially providing attackers with opportunities to escalate damage. Furthermore, many CPS devices manage essential physical functions, such as pressure monitoring in nuclear reactors, where improper reconfiguration or shutdown could result in disastrous outcomes. Additionally, while effective response strategies can be tailored to specific CPS configurations, the diversity of CPS applications, structures, and devices across various environments complicates the scalability of these solutions, making it difficult to generalize across deployments.

Some specific challenges are outlined below:

1A. Large, Powerful, High-fidelity Digital Twins: Designing and deploying advanced attacks and natural disasters that have not yet happened in the real-world requires a safe environment for testing. Digital twins can model advanced cyber attacks and other disruptions, and defenses, including AI-based, advanced social engineering etc.

Additionally, a significant challenge in advancing CPS security and resilience lies in the limited background knowledge and resources available for researchers to fully understand threats and conduct meaningful experimentation and evaluation. While cybersecurity researchers may be well-versed in identifying and categorizing various cyber threats, they often lack in-depth knowledge of the operational technology (OT) aspects specific to CPS, such as the control mechanisms, safety protocols, and physical processes that are critical to systems like power grids or industrial automation. This gap hinders their ability to accurately assess the real-world consequences of cyber threats on CPS and evaluate the practical effectiveness of defense mechanisms. Bridging this knowledge gap requires dedicated resources, including access to realistic CPS testbeds and cross-disciplinary collaboration, so that researchers can more effectively simulate, analyze, and respond to threats within these complex systems.

The power grid is an ideal example of a complex CPS that requires detailed testbeds for researchers to understand how to detect and prevent cyberattacks. Today's grid is becoming increasingly reliant on wide-scale, internet-connected distributed energy resources (DERs) such as photovoltaic (PV) systems, battery energy storage systems, and smart loads such as electric vehicles and HVAC systems. An attacker could gain access to one individual DER, or possibly gain control of many such systems by accessing a back end or cloud-based control system from an entity known as an aggregator that controls many DERs spread out over a wide geographic area. Testbeds are needed that can allow cybersecurity researchers to examine how such attacks can impact the real-time delivery of electricity. For example, a large-scale attack conducted in an appropriate manner on a large number of DERs could cause major disturbances and instability on the bulk power grid. In many cases, even an attack on millions of DERs may cause instability that is immediately detected by protection equipment. Attackers could begin to develop smart approaches, however, that detect the operation of protection systems and subsequently change the nature of the attack so as to cause cascading failures. Such an attack could take down large parts of the bulk power grid for days.

Developing testbeds of the scale described here requires work to develop tightly coupled co-simulation frameworks that can simultaneously simulate multiple interconnected networks. An effective test bed for the power grid, for example, should co-simulate electrical networks, the internet, and protection equipment. Implementing such a testbed at scale is an enormous challenge because of the vastness of the system.

1B. AI-based Attacks and Defenses for Critical Infrastructure: Although advancements in anomaly and intrusion detection have been notable, effective response to these attacks remains challenging with the rise of generative AI. The proliferation of quickly evolving Large Language Models could assist attackers in identifying more covert attack vectors and devising stealthier, more destructive attack strategies, complicating efforts to develop software patches, enforce resource isolation policies, and eliminate attack surfaces effectively. Existing security protocols, intrusion detection systems, and response frameworks are largely unprepared for the sophisticated, adaptive attacking strategies enabled by generative and other advanced AI technologies. This limitation leaves CPS particularly vulnerable, as AI-assisted attacks can dynamically chain together vulnerabilities and execute complex, multi-stage strategies that bypass conventional defenses.

The integration of AI, particularly *deepfakes*, into CPS presents a significant security risk. By manipulating sensor data or control signals, attackers could compromise the integrity and safety of critical infrastructure. This emerging threat necessitates a multi-faceted approach to cybersecurity, including robust authentication mechanisms, advanced threat detection techniques, and resilient system design principles.

1C. Behavioral Science for Critical Infrastructure Resilience: Verizon’s 2024 Data Breach Investigative Report indicated that 68% of breaches in the last year involved a non-malicious human element such as a person falling victim to a social engineering attack or making an error[1]. There are numerous major examples in recent years, including the Colonial Pipeline ransomware hack that took down the largest fuel pipeline in the US, which was caused by a compromised password and password reuse.

Multiple research gaps exist in this area. First, work is needed to understand how to simulate the human element in large-scale digital twins such as those described previously. Power grids with numerous internet-connected distributed energy resources (DERs), for example, constitute large scale CPS. An effective digital twin of the power grid should include models of human users. These models need to be incorporated into time-domain simulations and reflect how humans react to various social engineering attacks at scale, so that system planners can understand the impact that such attacks can have on the power grid.

Work is also needed to mitigate impacts from the human element in cyber-physical systems. Many industrial control systems used today, for example, rely on simple security standards and are installed on old desktop computers found in back rooms with access to a wider network. Examples include Building Automation Systems that control HVAC equipment in hospitals, data centers, and other critical facilities, or industrial control systems in water-treatment plants. Such equipment is particularly vulnerable, and the operators often have minimal training on concepts in cybersecurity. It's not uncommon, for example, to walk into a control room with a password included on a post-it note. Work is needed to train such operators, and similar work is needed as it relates to internet-connected distributed energy resources now finding their way into many homes.

1D. A Science of Cybersecurity: While cybersecurity is being studied extensively, it still lacks a truly scientific foundation. Unlike established sciences like physics or chemistry, cybersecurity often relies on ad-hoc solutions and reactive measures, and often ends up being a "cat-and-mouse" game between attackers and defenders. The rapid evolution of technology, adaptive adversaries, the complexity of systems today, the lack of standardized metrics, and the constant emergence of new threats, makes it challenging to establish a systematic, evidence-based approach.

1E. Under-Resourced Areas: The United States has a vast number of small organizations responsible for maintaining critical infrastructure. For example, 56% of the American landscape is served by electric cooperatives, many of which have a very small staff[2]. There are also numerous small & medium sized manufacturing facilities found in rural areas in the United States that rely heavily on automation but have very limited staff capabilities to prepare for cyberattacks.

[1] Verizon, https://www.verizon.com/business/resources/reports/dbir/
[2] https://www.electric.coop/electric-cooperative-fact-sheet

1F. Closing the Loop and Aging Infrastructure: CPS rely on the seamless integration of digital and physical components, forming a *closed-loop* system. These systems continuously monitor physical components, send data to digital twins for analysis, and use the results to adjust physical actuators.

However, the increasing sophistication of cyberattacks, particularly those leveraging AI-powered deepfakes, poses a significant threat to the security and reliability of CPS. Real-time communication and control are essential for effective response, but network latency and reliability issues can hinder timely actions. The increased connectivity between physical and digital components introduces new attack vectors—-malicious actors can exploit vulnerabilities in communication protocols, network infrastructure, or software to disrupt operations or compromise sensitive data. Data manipulation or unauthorized access can lead to incorrect decisions and compromised security. Physical access to critical infrastructure components can be exploited by attackers or be affected by natural disasters. Additionally, the supply chain for critical infrastructure components can be vulnerable to cyberattacks. Malicious actors may introduce compromised hardware or software, leading to security breaches.

Additionally, a big challenge in achieving “closing the loop” and rapid response is the typically aging infrastructure in most CPS.

2. Proposed Research Directions

To strengthen the robustness of essential infrastructure systems, research initiatives should bridge the fields of the critical infrastructure domain science science with cybersecurity and resilience innovations, targeting automatic and scalable solutions on both the discovery and reduction of system vulnerabilities. Based on current trends and emerging technologies, we propose the following research directions that NSF should consider prioritizing:

2A. Large, Powerful, High-fidelity Digital Twins that can model all aspects (e.g., the IT and OT sides, the humans involved, and the diverse and evolving nature) of the complex critical infrastructure ecosystem. These types of digital twins will allow modeling and testing of advanced natural disasters and cyber attacks and defenses, including AI-powered scenarios, advanced social engineering attacks etc.

Prominent research avenues should include developing advanced tools for simulating and modeling real-time interactions within critical infrastructures that account for the increasingly decentralized nature of critical infrastructure systems. New research programs should foster the collaboration between electrical engineers, scientists that study natural disasters (such as geologists, hydrologists etc.), cybersecurity specialists, high-performance computing experts, and data analysts. Also, programs should encourage the development of digital twins that enable researchers from diverse backgrounds to easily deploy and test their solutions, even if they have limited knowledge of other involved fields. Such a platform would lower the barriers to interdisciplinary collaboration, allowing researchers to validate and refine their methods in a broader context, thereby fostering knowledge sharing and technological integration across different domains.

Developing realistic representations of real CPS can be extremely complicated because of the vast nature of the interconnected networks. The power grid, for example, consists of many independently modeled electrical circuits running in conjunction with the internet. Constructing realistic models can be an extremely complex challenge requiring proprietary and sensitive data from utilities. In many cases, subject matter experts will need to get deeply involved to develop appropriate models. Advances in techniques such as federated learning could perhaps expedite the development of appropriate models.

2B. Advances in AI Techniques and Cyber Physical Resilience: Scalable and reliable vulnerability discovering and patching that aims to automate the existing security solutions with AI-empowered intelligent agents are an important area of research. These automation solutions will help overcome the scalability and cost issues of traditional security solutions that require specific expert knowledge or intervention. This helps the future autonomous systems such as vehicles and power grid to achieve self-sustainability against cyberattacks with acceptable costs.

Future research should also focus on integrating traditional security solutions and tools with the advanced reasoning capabilities of generative AI. Potential research tasks could range from developing security-focused AI agents to enhancing the reliability and robustness of these integrated systems. Achieving this goal requires approaches that address several key facets: (1) development of advanced AI models equipped with security-focused reasoning capabilities and up-to-date knowledge bases, enabling them to perform effectively in security tasks; (2) a comprehensive suite of security tools optimized for seamless integration with AI models, ensuring efficient information exchange and coordination; (3) improved organizational structures that incorporate AI-enabled agents into the existing workflows of critical infrastructures. New research programs should promote collaboration among experts in AI, cybersecurity, and operational sciences, encouraging applicants from each field to demonstrate how their proposed tasks will contribute to enhancing AI-enabled security solutions as a cohesive whole.

2C. Behavioral Science for Critical Infrastructure Resiliency: Developing effective digital twins such as the ones described above requires realistic models of human behaviors. Models that capture a range of possible responses to a social-engineering attack, for example, could be deployed as part of a multi-domain power-grid simulation. Work is needed to understand the statistical dependencies between certain attack responses and demographics, educational background, etc. Stochastic models could be deployed as part of a larger power-grid testbed. This is an area requiring fundamental work in behavioral science and close collaboration between researchers in different fields, including behavioral science, data science, and computation.

Work is also needed to understand how to better train OT infrastructure operators how to mitigate the impact of cyber breaches. Such operators can be operating critical systems, but do not commonly use even the simplest of best practices, including the use of effective passwords. Work is needed to understand how to best train and prepare these operators. Similar work is needed to address the homeowners that now have many DERs.

2D. A Science of Cybersecurity: Formal methods, memory-safe programming languages and other mathematical approaches to security are more critical than ever for CPS, where violation of policies can be life-threatening. By providing a rigorous approach to model, analyze, and verify systems, formal methods help identify and analyze threats, design secure systems, detect and respond to attacks, and develop *cybersecurity and resilience* metrics for a CPS[3].

While formal methods offer significant advantages, in order to effectively apply them, future research should tackle challenges such as complexity and scalability to handle the complexity of CPSs. Additionally, history has shown that formal methods can be heavy-weight, and difficult to adapt for non-experts. Future research should focus on lightweight formal methods, and ease of policy specification and enforcement for non-experts.

An important research area is developing effective techniques for combining formal methods with AI and behavior science of cybersecurity for critical infrastructure.

2E. Closing the Loop: Addressing these cybersecurity and resiliency challenges to “close the loop” in CPS requires a multifaceted approach, including robust cybersecurity measures, resilient network infrastructure, advanced threat detection and response capabilities, and a strong focus on human factors and supply chain security. Robust physical security measures, including access controls, surveillance, and intrusion detection systems, are essential to prevent unauthorized access and sabotage. Operators and engineers must be trained to respond effectively to cyber threats and avoid mistakes that could exacerbate security incidents.

Future CPS must incorporate resilient physical components that can actively respond to attacks in real-time, based on feedback from both the physical and digital realms. By intertwining these components with large, powerful, high-fidelity digital twins, we can create a more robust and secure CPS ecosystem.

[3] https://www.whitehouse.gov/wp-content/uploads/2024/02/Final-ONCD-Technical-Report.pdf

3. Education and Workforce Development

To ensure the sustainability and future growth of research in this field, NSF should also consider enhancing workforce development. We recommend:

- Supporting educational initiatives that train the next generation of scientists and engineers in the *intersection of AI and cybersecurity.*
- Support for long-term training of a cadre of personnel to maintain all elements of a CPS. This extends to those who will maintain elements on a daily basis (at the technician level and above), under normal circumstances and who must respond to unforeseen situations involving physical and cyber disruptions—-either through maintenance issues, natural events or as a result of malicious activities.
- Testbeds can serve a role in training and developing operational scenarios for normal operation as well as response to out of the ordinary conditions.
- Red Team exercises for CPS.
- Expanding funding for programs that promote diversity and inclusion in STEM, and specifically in cyber-physical systems resilience.
- Providing funding to develop unique cross-disciplinary educational programs designed to provide appropriate training to students in multiple fields such as power systems, behavioral science and computer science.
- Provide resources for academic institutions and industry partnerships for training talent to serve under-resourced smaller entities (such as rural electric cooperatives) who are critical to America’s national security.

4. Collaboration and Infrastructure

- DOD and civil systems can benefit from the cross-pollination of approaches. For example, as AI becomes an element of the Tactical Microgrid concept and that construct infuses the domain of fixed facilities (as US Army DEVCOM has identified as a desired objective) there will be a growing need for capabilities to assess, validate, and train.
- Creating mechanisms to facilitate multi-institutional collaborations, including shared research infrastructure.
- Collaboration between researchers across institutions and industries is crucial for advancing this field.
- Increasing funding for the construction (hardware, software, expertise) of large-scale experimental testbeds and computational infrastructure to support advanced research in CPS.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Mitchell Berger

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Mitchell Berger
[redacted]
October 20, 2024

To: Melissa Cornelius, NITRD, [redacted]

Re: **Networking and Information Technology Research and Development Request for Information (RFI) on a National Plan for Cyber-Physical Systems Resilience, https://www.regulations.gov/document/NSF_FRDOC_0001-3403/comment**

Note: This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

Dear Ms. Cornelius: In response to NITRD's and NSF's above RFI, I write to emphasize the importance of focusing on cyber-physical systems (CPS) resilience in the context of health and healthcare. The PCAST plan referenced in the RFI, while including participants from the health sector, does not particularly emphasize health care needs or challenges.[1] On the other hand, the Department of Health & Human Services' Healthcare and Public Health Sector-Specific Cyber Performance Goals have limited discussion of CPS.[2]

Research needs concerning medical CPS have been noted.[3] One group of authors points out that "There is [….] an urgent need for an empirical study on MCPS and the nature of the relationship between the key characteristics of healthcare service delivery. The studies should also explore the use of MCPS in preventative, curative, palliative and rehabilitation healthcare services in terms of SWOT analysis."[4] Within health care, the Internet of Things (IoT) can support for patient monitoring and treatment.[5] Such devices may be overlooked in traditional vulnerability assessments.[6] In some cases, medical devices and other technologies have been accessible from the public internet.[7] 0

One company studying IoT at 300 hospitals notes that "53% of connected medical and other IoT devices in hospitals have a known critical vulnerability, including a third of the bedside devices

[1] https://www.whitehouse.gov/pcast/briefing-room/2024/02/27/pcast-releases-report-on-strategy-for-cyber-physical-resilience/; Xu H, Yu W, Griffith D, Golmie N. A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective. IEEE Access. 2018;6:10.1109/access.2018.2884906. doi: 10.1109/access.2018.2884906. PMID: 35531371; PMCID: PMC9074819.

[2] https://hhscyber.hhs.gov/documents/cybersecurity-performance-goals.pdf

[3] Sony, M., Antony, J. and McDermott, O. (2022), "The impact of medical cyber–physical systems on healthcare service delivery", The TQM Journal, Vol. 34 No. 7, pp. 73-93. https://doi.org/10.1108/TQM-01-2022-0005

[4] For more information on SWOT see https://www.mindtools.com/amtbj63/swot-analysis; https://ndupress.ndu.edu/Media/News/News-Article-View/article/2053087/cyber-physical-systems-the-coming-singularity/

[5] Chunyan Li, Jiaji Wang, Shuihua Wang, Yudong Zhang, A review of IoT applications in healthcare, Neurocomputing, 2024: 565: 127017, https://doi.org/10.1016/j.neucom.2023.127017; Verma R. Smart City Healthcare Cyber Physical System: Characteristics, Technologies and Challenges. Wirel Pers Commun. 2022;122(2):1413-1433. doi: 10.1007/s11277-021-08955-6.

[6] https://claroty.com/blog/10-examples-of-cyber-physical-systems

[7] https://medcitynews.com/2024/10/cybersecurity-healthcare-2/; https://www.cdotrends.com/story/17594/iot-security-giving-healthcare-heart-attacks

that patients most depend on for optimal health outcomes."[8] Devices using outdated operating systems, including IV pumps and nursing call systems, may place patients at risk.[9]

CPS also is beginning to support within health care use of robotics, enhanced data management (e.g., through use of blockchain technologies) and personalized care.[10]

In addition to CPS within the health sector itself, health care agencies and organizations also are reliant on other critical infrastructure sectors such as food and water and transportation and electricity and vice versa as many of these sectors would rely on the health care sector in an emergency.[11]

NITRD, NSF and partners should support further research into how CPS is used within health care and interdependencies within health care and other sectors using CPS. Privacy and confidentiality also should be further understood and emphasized.[12] NITRD and its partners also should research how best to obtain input from members of the public and end-users. As CPS capabilities evolve, so too will ethical and policy questions about use of these systems and the data they collect become more complex. Obtaining public input on an ongoing basis will be essential.[13] There also is a need to offer objective, comprehensive education about these topics to state, local, tribal and territorial policymakers and their staff.

I appreciate your consideration of this input.

Sincerely,

Mitchell Berger

Note: Please note that I am submitting these suggestions in my personal/private capacity and not on behalf of any organization or agency. The views expressed are mine only and should not be imputed to other individuals nor to any public or private entity.

[8] The State of IoMT Device Security, Cynerio, https://www.cynerio.com/landing-pages/the-state-of-healthcare-iot-device-security-2022; https://www.cynerio.com/blog/visibility-is-not-enough-key-takeaways-from-cynerios-2022-state-of-healthcare-iot-device-security-report

[9] Id., See also, https://healthitsecurity.com/news/1-in-5-connected-medical-devices-run...1

[10] H. M. Khater et al., "Empowering Healthcare With Cyber-Physical System—A Systematic Literature Review," in IEEE Access, vol. 12, pp. 83952-83993, 2024, doi: 10.1109/ACCESS.2024.3407376; https://www.nationalacademies.org/news/2023/11/the-rise-of-cyber-physical-systems

[11] https://www.naccho.org/blog/articles/the-department-of-homeland-security-sector-risk-snapshots-released

[12] https://pages.nist.gov/cpspwg/; Differential Privacy Techniques for Cyber Physical Systems: A Survey. MU Hassan, MH Rehmani, J Chen. IEEE Communications Surveys & Tutorials 22 (1), 746-789.

[13] See e.g., https://ecastnetwork.org/;https://participedia.net/

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

MITRE

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

MITRE's Response to the NSTC RFI on Cyber-Physical Research

October 25, 2024

For additional information about this response, please contact:

Duane Blackburn
Center for Data-Driven Policy
The MITRE Corporation
[redacted]

[redacted]

"This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution."

About MITRE

MITRE is a not-for-profit company that works in the public interest to tackle difficult problems that challenge the safety, stability, security, and well-being of our nation. We operate multiple federally funded research and development centers (FFRDCs), participate in public-private partnerships across national security and civilian agency missions, and maintain an independent technology research program in areas such as artificial intelligence, intuitive data science, quantum information science, health informatics, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience. MITRE's ~10,000 employees work in the public interest to solve problems for a safer world, with scientific integrity being fundamental to our existence. We are prohibited from lobbying, do not develop or sell products, have no owners or shareholders, and do not compete with industry—allowing MITRE's efforts to be truly objective and data-driven. Our multidisciplinary teams (including engineers, scientists, data analysts, organizational change specialists, policy professionals, and more) are thus free to dig into problems from all angles, with no political or commercial pressures to influence our decision making, technical findings, or policy recommendations.

MITRE's extensive experience in cyber-physical research is rooted in its unique position as a trusted advisor to the government, enabling the organization to address complex challenges in cyber resilience and security. MITRE's work in this domain includes developing advanced analytics and prototype solutions that enhance both cyber defense and offense capabilities. By leveraging its expertise in areas such as machine learning, operations research, and game theory, MITRE has made significant strides in understanding and mitigating cyber threats to critical infrastructure, including transportation systems and military networks.

MITRE's approach emphasizes integrating cyber and physical systems to create resilient and secure environments. This includes applying models to identify vulnerabilities and unifying cyber and electronic operations. MITRE demonstrates its commitment to advancing the state of the art in cyber-physical systems through collaboration with government, industry, and academia, ensuring that solutions are both innovative and practical. Operating without commercial or political pressures, MITRE provides objective, data-driven insights critical to shaping national cyber-physical research strategies.

Overarching Recommendations

Building on MITRE's extensive experience and commitment to addressing complex national challenges, we present a set of overarching recommendations designed to enhance the resilience and security of cyber-physical systems. These recommendations aim to integrate innovative research, foster collaboration, and drive strategic advancements across sectors. By focusing on resilience by design, predictive defense technologies, and cross-sector partnerships, we can develop comprehensive strategies that safeguard critical infrastructure and ensure a safer, more secure future.

Integrate Cyber-Physical Security into a Unified Research Framework. Integrating cyber-physical security into a unified research framework is essential for addressing the complex threats facing critical infrastructure. By combining cybersecurity, physical security, and operational resilience, the nation can develop comprehensive protection strategies that address vulnerabilities across all dimensions. This approach fosters collaboration between experts, creating holistic solutions that anticipate and mitigate a wide range of threats. Research should focus on advanced threat detection and response mechanisms that operate seamlessly across domains, leveraging technologies like artificial intelligence (AI) and machine learning (ML). This integrated framework will inform the development of cohesive security strategies, ensuring robust protection against evolving threats.

Focus on Resilience, Composability, and Secure Design. Encourage research and development (R&D) initiatives that embed resilience principles into the design and operational phases of cyber-physical systems. Research should explore adaptive and composable systems capable of recovery across various threat scenarios. Additionally, ensure that systems-of-systems are secure by design, promoting integrated and holistic security measures throughout their lifecycle.

Advance Research in Predictive and Proactive Defense Technologies. Support research into AI, ML, and advanced modeling and simulation to enhance predictive threat analysis and real-time response capabilities. These technologies can significantly improve system resilience.

Foster Cross-Sector Research Collaboration. Encourage collaborative research efforts between government, industry, and academia to share insights and innovations. This collaboration will drive the development of resilient systems and enhance the nation's collective security posture.

Integrate Public Perceptions into Resilience Strategies. A recent MITRE-Harris poll[1] reveals that while the American public recognizes the critical role of infrastructure security, there are significant concerns about recovery capabilities and differing views on responsibility. With 29% believing recovery is solely a federal duty and 49% seeing it as a shared responsibility among various government levels and private operators, these perceptions are crucial for national-level planning. By incorporating public expectations into the national cyber-physical research strategy, we can build trust and cooperation, ensuring policies are both technically robust and socially supported. This alignment is vital for effective implementation and public buy-in, ultimately enhancing national resilience and security.

Inputs Requested in RFI

1. Threat-agnostic approaches for resilience are of special interest. As part of the input, we are primarily concerned with the ability of cyber-physical systems to recover and adapt while ability to withstand may be already covered in the current risk assessment and management efforts. We are particularly interested in how resilience by design or resilience by intervention can prepare for recovery and adaptation in different threat scenarios as well as in threat-agnostic situations.

Develop and Direct Research Initiatives to Establish Clear Metrics for Assessing and Prioritizing the Resilience of Cyber-Physical Systems (CPS). This research should focus on defining key performance indicators that measure system robustness, adaptability, and recovery time, providing a common framework for evaluating and enhancing the security and reliability of

[1] MITRE-Harris Poll Finds U.S. Public Is Worried about the Security of Our Critical Infrastructure. 2024. MITRE, https://www.mitre.org/news-insights/news-release/mitre-harris-poll-finds-us-public-worried-about-security-our-critical, last accessed October 15, 2024.

critical infrastructure. By creating a standardized approach to operationally prioritize, test, and evaluate CPS survivability, these metrics will not only guide the design of more resilient systems but also inform discussions on the relative resiliency of existing technologies, ultimately leading to improved strategies for mitigating disruptions.

Enhance Data Integration for Real-Time Modeling and Simulation. To maximize the effectiveness of modeling and simulation (M&S) in cyber threat intelligence, research should aim to improve the integration of high-quality, real-world, and real-time data into simulation environments. This includes developing methods to seamlessly incorporate diverse data sources, ensuring simulations accurately reflect current system states and potential adversary actions. By encompassing multi-domain threat modeling, such as large-scale cyber attacks following significant weather events, enhanced data integration will lead to more precise analyses, enabling better-informed decisions and improved system resilience.

Establish Methods for Prioritizing Investments into Resilient Cyber-Physical Systems. Research should focus on expanding the impact of design principles that integrate security considerations early in the development of CPS, such as Cyber-informed Engineering, Secure by Design, and Zero Trust architectures. By adopting a mindset that assumes systems are compromised, engineers can implement redundancy, diversity, and failsafes to limit the impact of cyber attacks. However, implementation of these principles after systems have been designed or deployed remains prohibitively expensive in many cases. Instead, research should focus on developing strategies and methods to prioritize security enhancements and mitigation techniques. Prioritization should consider threat actor capabilities, intent, and future growth *in addition* to the impact or consequence of a cyber-induced adverse event. This proactive, but threat-actor-agnostic, approach will enhance system resilience, making it more difficult for adversaries to exploit vulnerabilities, especially as emerging technologies like AI and ML are increasingly adopted.

Conduct Research to Evaluate and Streamline Security Regulations for Critical Infrastructure. Research should focus on evaluating the effectiveness of current security regulations and identifying outdated or ineffective requirements. This involves developing methodologies to assess regulatory impact and streamline guidance to align with modern threat landscapes. By reducing unnecessary burdens, this research will help critical infrastructure operators optimize resource allocation and enhance overall security posture.

Recommend Best Practices for Operational Continuity and Resilience. To enhance coverage against cyber threats, weather events, and system failures, research should focus on developing best practices for operational continuity and resilience. This involves identifying strategies for redundant monitoring, manual overrides, and built-in safety thresholds, particularly in remote management systems like Industrial Control Systems (ICS). By addressing these areas, we can mitigate impacts and ensure robust protection for cyber-physical systems, reducing vulnerabilities and enhancing overall system resilience.

2. Examples of domains and application of interest include but are not limited to critical infrastructure and systems for energy, transportation, medical, agriculture, water, space, manufacturing, and other R&D topic areas in which the strategic plan should focus, as well as details that should be considered when/if the topic area is elaborated in the strategic plan.

Align Research Along CISA's National Critical Functions. Aligning this cyber-physical research strategy with DHS/CISA's National Critical Functions (NCFs) is strategically beneficial and represents sound policy. The ongoing cataloging and updating of the NCFs allows for prioritized response and risk management, evolving from an entity-driven approach to one focused on functional outcomes. This alignment ensures endeavors are integrative, maximizing impact across sectors.

- **Understanding Interdependencies and Risks**: The complex interdependencies within and across U.S. critical infrastructure heighten the risks to continuous operation of the NCFs. Recognizing these interdependencies is essential for effective prioritization and risk management.
- **Strategic Risk Management**: The NCF set is organized into four areas: Connections, Distribution, Management, and Supplies. While risk management has progressed beyond entity-driven hardening, national-level impacts can still arise from cascading system-level threats, such as sophisticated cyber attacks, CPS compromise, natural disasters, or systemic failures. Further expanding tool capabilities and conducting advanced tabletop exercises will facilitate broader understanding of the interconnectedness, strengths, and weak spots of the NCF areas.
- **Innovative Approaches and Tools**: Expanding NCF area-level actions and cross-area strategies represents the next evolution of risk management. The STAR (Suite of Tools for the Analysis of Risk), hosted on CISA's Mission Critical Test Environment (MCTE), connects the 55 NCFs to assets, sectors, and responsible owners, with STAR v2 enhancing capabilities through advanced tools and datasets. MITRE can leverage numerous existing and future studies, datasets, and technologies to continue building on the early foundation of STAR.

Incorporate New Risk Mitigation Strategies:

- **Avoidance and Mitigation**: Risks can be avoided, mitigated, transferred, or accepted, but effective management requires a comprehensive understanding of threat scenarios. While individual risks may be inevitable, we must proactively address potential cascading failures. Doing so may require new real-time operational models, dashboards, and reporting mechanisms.
- **Decoupling and Autonomy**: Eliminating risks involves decoupling critical systems or enabling continuous operation through technological innovations, such as autonomous systems with AI-enhanced offline control. Digital twinning and enhancing localized capabilities can be researched and validated.
- **Real-Time Mitigation**: Implementing near-real-time stop-gap actions to control problems and enhance human decision making is crucial. Exploring extensible control mechanisms across disparate systems can enhance resilience. This will require new modeling, data categorization, and reporting across multiple sectors.

- **Risk Sharing**: Increasing interconnectedness through actionable intelligence and legal protections can enhance redundancy and defense-in-depth, allowing for flexible emergency functions across technologies and networks.

These strategies should be explored through technical studies, detailed modeling, and pilots to enhance and/or complement the STAR program. Grounding the research strategy in CISA's framework ensures a robust approach to safeguarding critical infrastructure, making it essential for a comprehensive national strategy.

Innovate Tools for Dynamic Attack Surface Management in Critical Infrastructure. As critical infrastructure becomes increasingly digital and interconnected, managing the attack surface is essential to minimizing vulnerabilities. Research should aim to develop advanced tools and methodologies that dynamically assess and reduce the attack surface, balancing the need for operational visibility with security. By innovating in this area, we can better protect infrastructure from cyber threats while maintaining the efficiency and effectiveness of operations.

Develop High-Fidelity Modeling and Simulation Frameworks for Critical Infrastructure. Research should focus on creating advanced M&S frameworks that leverage digital twin technology and virtualization to provide realistic environments for testing and evaluation of critical infrastructure systems. These frameworks can help asset owners and operators understand vulnerabilities and optimize resource allocation for security. By reducing costs and improving accessibility, these M&S tools can enhance decision making and resilience at both regional and national levels.

Additional Examples of Domains and Applications of Interest (Non-exhaustive):

- **Smart Transportation, Autonomous Vehicles, Vehicle-to-Infrastructure (V2X) Security**: Secure, safe, and resilient transportation infrastructure is critical to the world economy. Vehicle cybersecurity still needs improvement, and relatively little focus has been placed on ensuring the overall transportation infrastructure is resilient. The vehicle community has largely relied on manufacturers to set and meet their own security standards, and is often limited by lack of data on realistic threats and resilience metrics. In addition, vehicle manufacturers are not modeling the V2X environment or self-nominating security requirements to ensure the overall infrastructure is secure.
- **Energy Infrastructure**: Secure and resilient energy infrastructure is vital for economic stability and national security. The power grid faces increasing cyber threats, yet there is a need for more comprehensive frameworks to enhance its resilience. The energy sector often lacks real-time threat detection and response capabilities, which are crucial for preventing disruptions. Additionally, there is insufficient data on potential threats, resilience metrics, and dynamic interdependencies on the sector, limiting the ability to model and mitigate risks effectively. Collaborative efforts between stakeholders are essential to establish robust security standards and ensure the grid's integrity.
- **Medical Devices and Healthcare Systems**: The cybersecurity of medical devices and healthcare systems is critical to patient safety and data protection. As connectivity increases, these systems become more vulnerable to cyber threats. The healthcare community often relies on manufacturers to self-regulate security standards, which can lead to inconsistencies. There is a pressing need for comprehensive guidelines that address authentication, data encryption, and regular security updates. Ensuring the security of

medical devices, systems, and networks requires a coordinated approach to protect sensitive patient data and maintain device functionality.

- **Water Systems**: Protecting water supply systems from cyber threats is essential for public health and safety. Water treatment facilities and distribution networks are increasingly targeted, yet there are continued inconsistencies in the implementation of security practices and standards. The water sector often struggles with risk assessment and incident response capabilities, which are crucial for safeguarding against potential attacks. Developing robust security measures and conducting regular vulnerability assessments are necessary to ensure the integrity and reliability of water systems.
- **Space Systems**: The security of space systems is critical as reliance on satellites for communication and navigation grows. Space assets face unique cyber and physical threats, yet there is limited focus on comprehensive protection strategies. The space industry also provides additive and alternate communication paths for numerous critical functions and thus requires additional risk mitigation strategies and enhanced security measures for the transport of sensitive data. Collaborative efforts are needed to develop strategies that protect these assets, ensuring the reliability and continuity of space-based services.
- **Manufacturing**: Manufacturing processes depend heavily on ICS, which are susceptible to cyber-physical threats. The sector often lacks consistent implementation of security best practices, industry standards, and cybersecurity maturity models to measure against, leaving ICS vulnerable to attacks that can disrupt production. There is a need for best practices in network segmentation, access control, and incident response to protect manufacturing operations. Establishing comprehensive security standards and conducting regular assessments are crucial for safeguarding the integrity of manufacturing systems.

3. Other Inputs. (This RFI seeks input to shape a whole-of-government effort on research and development that will strengthen cyber-physical resilience. In the context of this RFI, we refer to threats to include cybersecurity, physical, natural disasters including extreme weather events or other hazards such as earthquakes, and the potential for adversary use of AI to disrupt systems as well as deceive human operators of critical infrastructure systems.)

Further Consider the PCAST Recommendation for a National Critical Infrastructure Observatory. We encourage further consideration of the President's Council of Advisors on Science and Technology (PCAST) recommendation[2] to establish a National Critical Infrastructure Observatory aimed at enhancing resilience across U.S. critical infrastructure. This observatory would serve as a "lived-in" testbed for real-world applications, focusing on a range of threats, including natural disasters and cyber risks.

- **Testbed Selection**: We recommend selecting representative municipalities that face typical critical infrastructure challenges and common external threats, such as weather-related events. Cities like Houston and Norfolk, known for their vulnerability to disasters and their

[2] Strategy for Cyber-Physical Resilience: Fortifying Our Critical Infrastructure for a Digital World. 2024. President's Council of Advisors on Science and Technology, https://www.whitehouse.gov/wp-content/uploads/2024/02/PCAST_Cyber-Physical-Resilience-Report_Feb2024.pdf.

critical infrastructure needs related to the Defense Industrial Base, would be ideal candidates for this initiative.

- **Public-Private Collaboration**: A successful observatory must prioritize public-private partnerships, bringing together government agencies, private sector stakeholders, and academic institutions. This collaboration will be essential for understanding diverse challenges and developing comprehensive solutions, ensuring that the observatory serves as a vital connector across sectors.
- **Focus on Grand Challenges**: The observatory should address grand challenges, such as leveraging AI and advanced technologies to enhance resilience. By focusing on these challenges in a real-world context, the observatory can generate insights and frameworks that effectively improve national cyber-physical resilience.
- **Community Engagement**: While there may be concerns from residents about participating in observatory municipalities, it's crucial to highlight the benefits of engagement. Living within these environments will facilitate practical learning and preparedness, ultimately enhancing community resilience.

By utilizing the observatory as a connector and testing ground for innovative solutions, we can ensure that research translates into meaningful advancements in infrastructure security and resilience across the country.

Advance Predictive Cyber Threat Intelligence Methodologies. To enhance the utility of cyber threat intelligence in building resilient systems, research should focus on developing predictive analysis methodologies that anticipate future threats rather than solely relying on retrospective analysis. This involves creating models that incorporate diverse data sources and leverage machine learning to forecast potential adversary actions. By advancing predictive capabilities, organizations can proactively address vulnerabilities and better allocate resources, ultimately improving system resilience and reducing risk.

Integrate Engineering Expertise into Cyber Threat Intelligence. To improve the accuracy and applicability of threat intelligence, research should prioritize the integration of engineering expertise into the intelligence production process. This involves developing frameworks that facilitate collaboration between cyber intelligence analysts and engineering subject matter experts, ensuring a comprehensive understanding of how critical infrastructure systems are designed and operated. By bridging this gap, threat intelligence can more effectively inform risk assessments and resilience strategies, leading to more robust protection of critical assets.

Investigate Defensive AI and ML Applications for Cyber Resilience. To counteract the adversarial use of AI and ML, research should explore defensive applications of these technologies to enhance cyber resilience. This includes developing AI-driven tools for automated vulnerability detection, defense design and optimization, threat prediction, and real-time response. By leveraging AI and ML defensively, organizations can stay ahead of adversaries, improving their ability to protect critical infrastructure from sophisticated cyber threats.

Direct Science and Technology Funding to Investigate "Break Glass" and Other Crisis Technologies. As demonstrated during the "Shields Up" U.S. government initiative ahead of the Russian invasion of Ukraine, there is value in increasing security postures in times of crisis. Unfortunately, maintaining a more robust security posture can strain operations, putting substantial burden on staff and systems. Similarly, recovery actions following a successful breach or attack can also strain operations, and past events have demonstrated that organizations that respond quickly in the event of cyber attack can reduce overall damage.[3] Some past research has demonstrated the utility of these crisis technologies,[4] but additional R&D direction and focus would accelerate development and adoption. MITRE recommends that the research strategy allocate more R&D resources to enhance the resiliency of CPS against adversary attack.

[3] The 2015 cyber attack in Ukraine against three electricity distribution companies demonstrated the benefits associated with fast responses. In that case, the utility that quickly severed its virtual private network connections had the least interruptions in power and the fastest restoration times.

[4] One such example is Idaho National Laboratory's Constrained Communications Cyber Device, which can be used to restrict communications with protective relays in the event that an imminent attack on electric grid operations is suspected. (Additional information is available at https://inl.gov/national-security/prpc/.)

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Pacific Northwest National Laboratory

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to 89 FR 78915

Pacific Northwest National Laboratory

October 2024

Thomas Edgar
David Manz
Elena Peterson
Thomas Carroll
Jessica Smith
Marissa Decillis
Veronica Adetola
Michael Henson
Lori Ross O'Neil

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory (PNNL) has extensive capabilities and experience in cyber-physical security and resilience. We develop integrated security solutions that address both cyber and physical vulnerabilities in critical infrastructure systems like energy, transportation, and water sectors. Engaged in advanced research and development, PNNL works to identify and mitigate emerging riks associated with the convergence of cyber and physical domains and create cutting edge solutions like applying zero trust concepts into operational technology environments and utilizing deception as an easy-to-use industry detection technique. Our specialization in threat analysis and detection involves the use of machine learning and AI to enhance the identification and response to potential attacks.

PNNL is also focused on resilience engineering through investments like the Resilience Through Data-Driven, Intelligently Designed Control (RD2C) Initiative, working to enhance the ability of infrastructure systems to recover quickly and maintain operations following disruptions. Using sophisticated modeling and simulation tools, it is possible assess the impact of cyber-physical threats and test the effectiveness of security measures in controlled environments. The laboratory collaborates with government, industry, and academic partners to develop standards, share best practices, and advance the field of cyber-physical security. Additionally, PNNL offers training and educational programs that prepare organizations and individuals to meet modern cyber-physical security challenges.

PNNL's Response: A National Plan for Cyber-Physical Systems Resilience

Cyber resilience refers to an organization's ability to continuously deliver the intended outcome, despite adverse cyber events. It encompasses the abilities to anticipate, withstand, recover from, and evolve to improve capabilities against future cyber threats. The most common definitions of cyber resilience generally include the following components:

- Preparation: The capacity to predict and prevent cyber threats. This involves conducting risk assessments, threat modeling, and implementing proactive defense strategies.
- Resistance: The ability to resist and endure when a cyber incident occurs without significant disruption. This includes having strong preventive measures like firewalls, intrusion detection systems, and robust security protocols in place.
- Recovery: The speed and effectiveness with which an organization can recover from a cyber incident. Effective disaster recovery plans, data backup solutions, and incident response teams are critical to this aspect.
- Adaptation: The ability to learn from cyber incidents and evolve to be better prepared for future threats. This includes analyzing incident reports, updating security policies, and investing in continuous training and improvement.

The three commonly used definitions of cyber resilience are from NIST, DoD, and DHS.

National Institute of Standards and Technology (NIST)[1]

NIST describes cyber resilience as the ability to anticipate, withstand, recover from, and adapt to adverse conditions, stresses, attacks, or compromises on information systems. Key Elements of their definition includes anticipate (Forecasting potential threats and vulnerabilities), withstand (Robust measures to prevent and resist attacks), recover (Efficient recovery mechanisms post-incident), adapt (Continuous learning and evolving security practices). The definition focuses on balancing preventive, detective, and

[1] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-160v1.pdf

corrective controls and emphasizing on a holistic approach encompassing technology, processes, and people.

Department of Defense (DoD)[1]

The DoD defines cyber resilience as the ability of a system or organization to resist, absorb, recover from, or successfully adapt to adverse occurrences that could cause harm, such as cyber-attacks. Key Elements of their definition includes resisting (Defend against attacks), absorbing (Minimize the impact of attacks), recovering (Restore normal operations swiftly), and adapting (Enhance defenses based on past experiences). The definition has a strong emphasis on mission assurance and operational continuity and incorporates a warfare context in which cyber resilience is crucial to maintaining mission effectiveness under adversarial conditions.

Department of Homeland Security (DHS)[2]

According to DHS, cyber resilience refers to the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions due to deliberate attacks, accidents, or naturally occurring threats or incidents. Key Elements of their definition includes preparing (Strategic planning and readiness), adapting (Flexibility to handle evolving threats), withstanding (Measures to resist damage), and recovering (Rapid restoration of services). The definition includes a broader scope including deliberate, accidental, and natural disruptions, emphasizes critical infrastructure protection and public-private partnerships and focuses on national security and public safety.

These definitions stress preparation, resistance, recovery, and adaptation as key elements. Each emphasizes the importance of minimizing disruption to crucial services or operations. Overall, cyber resilience strives to ensure the operational continuity and security of an organization's critical services in the face of ongoing cyber challenges and enhancing the ability to develop better defenses.

Digital systems resilience can be robustly achieved through a synergistic approach encompassing both strategic planning and tactical operations. From a strategic perspective, resilience begins with the intentional design and configuration of digital systems to possess inherent resilient attributes. This involves developing more distributed, fault-tolerant system architectures that eliminate single points of failure and ensure continuity of service. Integrating Zero Trust principles into these systems—where no entity is trusted by default—further solidifies the security posture by meticulously managing access controls and continuously verifying every request. Understanding and documenting the base vulnerability states of these systems through comprehensive risk assessments and vulnerability scans is also crucial. Such foundational knowledge enables organizations to anticipate and mitigate potential threats before they materialize, fundamentally strengthening the resilience of the digital infrastructure.

On the tactical and operational front, resilience is further ensured by actively monitoring and responding to threat activities through proactive and adaptive defenses. This involves the deployment of advanced monitoring tools and systems that provide real-time visibility into network activities, detect anomalies, and identify potential security incidents. Proactive defense mechanisms, such as threat intelligence-driven deceptions and automated incident detection and response systems, are essential in identifying and neutralizing threats before they cause significant damage. Adaptive defenses, which dynamically adjust to evolving threats, ensure that security measures remain effective in the face of new and sophisticated attack techniques. By coupling these advanced controls with continuous performance measurement and monitoring, organizations can not only respond to threats more efficiently but also iteratively improve their

[1] https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/302040p.pdf

[2] https://www.dhs.gov/publication/2010-risk-lexicon

resilience strategies. This dual approach of strategic planning and tactical execution creates a resilient digital ecosystem capable of withstanding, recovering from, and adapting to an array of adverse conditions and cyber threats.

A resilient response is multifaceted, achieved by integrating resilience across various layers of an organization's infrastructure. These layers include supply chain, control system design and operation, communication network design and operation, and operator response support.

Resilience through Supply Chain

Supply chain resilience is a critical component of an overall cyber resilience strategy. It involves ensuring the integrity, security, and robustness of all components and services essential for operations. By integrating resilience into the supply chain, organizations can better prepare for and mitigate disruptions, whether caused by cyber-attacks, natural disasters, or other adverse events. Key activities in this regard include collecting comprehensive build materials, tracking vulnerabilities, and establishing a national vulnerability notification system. Here's a detailed overview:

1. Collecting Bills of Material for Software, Hardware, and Firmware

A Bill of Materials (BoM) is a detailed list of all components, including software, hardware, and firmware, used in an organization's IT infrastructure. Enables precise tracking of components and their interdependencies. Facilitates efficient vulnerability management and patching processes. Enhances transparency and accountability in the supply chain.

Strategic Activities:

- Inventory Management: Maintain an up-to-date inventory of all components along with their versions, manufacturers, and dependencies.
- Automated Tools: Employ automated tools to scan and document all software libraries, hardware parts, and firmware versions.
- Supplier Collaboration: Work closely with suppliers to obtain accurate and comprehensive BoMs for all products.

2. Tracking and Mapping Vulnerabilities

Continuously monitoring and identifying vulnerabilities in software, hardware, and firmware components is crucial for proactive defense. Reduces the attack surface by proactively addressing vulnerabilities. It is not currently possible for cyber-physical system operators to determine their full cyber risk exposure for the equipment they have procured and operate. Creating a method and framework to better inform operators how newly discovered vulnerabilities relate to their operations helps them more quickly mitigate the problem, reducing the window for threat exploitation.

Strategic Activities:

- Vulnerability Databases: Utilize databases like NIST's National Vulnerability Database (NVD) to stay informed about known vulnerabilities.
- Automated Scanning: Implement automated vulnerability scanning tools to identify and assess vulnerabilities in real-time.
- Risk Assessment: Conduct regular risk assessments to prioritize vulnerabilities based on their potential impact and likelihood of exploitation.
- Patch Management: Develop and enforce a robust patch management process to promptly address and mitigate identified vulnerabilities.

3. Creating a National Vulnerability Notification System

Establishing a centralized system at the national level to inform organizations about new vulnerabilities and their potential impact on their environments. Provides organizations with timely and relevant information to address vulnerabilities. Enhances national cybersecurity posture by ensuring widespread awareness and coordinated response to emerging threats. Reduces the likelihood of widespread exploitation of vulnerabilities by enabling proactive defenses.

Strategic Activities:

- Centralized Repository: Create a national repository that aggregates vulnerability data from various trusted sources.
- Real-Time Alerts: Develop a system to disseminate real-time alerts and notifications about new vulnerabilities to relevant stakeholders.
- User Registration: Allow organizations to register and specify their environments so they receive tailored notifications based on their specific configurations.
- Collaboration and Coordination: Facilitate collaboration between government agencies, private sector entities, and security researchers to enrich the repository with comprehensive, timely information.

Resilience through Control

Ensuring resilience through control design and operation is crucial for maintaining the functionality and security of critical infrastructure systems, even in the face of cyber threats. This strategy focuses on designing systems with resilience, using simulations to understand risks, implementing new control frameworks for distributed infrastructure, and developing metrics to enhance resilience management. Key elements of the strategy include:

1. Resilience by Design Using Simulations

Designing systems with resilience from the start and using simulations to understand cyber risks, identify weaknesses, and optimize control frameworks and mitigation strategies. Provides a proactive approach to cybersecurity by understanding and addressing potential risks before they materialize. Optimizes resource allocation by identifying the most effective controls and mitigations. Enhances the robustness and reliability of control systems.

Strategic Activities:

- Threat Modeling: Develop threat models to identify potential attack vectors and vulnerabilities.
- Co-Simulation Tools: Utilize advanced simulation tools to simulate cyber-attacks and their impact on systems. Exploring across both cyber and physical models is crucial to understanding and evaluating the emergent behavior and risks as the two become increasing intertwined. Co-simulation and integrated modeling tools are crucial to answer the pressing and upcoming securing questions.
- Scenario Analysis: Conduct scenario-based analyses to evaluate the effectiveness of different control frameworks and mitigation strategies. It is important to understand both the highest risk/probable threat scenarios against an existing system and the most reoccurring targets that cause physical impacts to better inform decision makers on how and where to prioritize resilient planning.
- Optimal Control Selection: Use co-design approaches with simulation results to select optimal control frameworks and cyber mitigations tailored to specific risks and operational requirements. Since

resources are limited, it is crucial to create tools that help asset owners prioritize and optimize investments for maximum risk mitigation benefit.

2. Implementing New Control Frameworks for Distributed Infrastructure

Adopting and implementing control frameworks and approaches that cater to the more distributed nature of modern critical infrastructure systems. Increases system resilience by reducing single points of failure. Enhances the ability to localize and contain disruptions. Supports scalability and flexibility in managing critical infrastructure.

Strategic Activities:

- Decentralized Control Systems: Develop and implement decentralized control systems that can operate independently and collaboratively.
- Distributed Architecture: Design systems with a distributed architecture to enhance redundancy and fault tolerance.
- Edge Computing: Utilize edge computing to process data locally, reducing reliance on centralized systems and improving response times.

3. Cyber Attack-Aware Systems Designed to Fight Through Adverse Conditions

Developing systems that are aware of cyber-attacks and designed to continue operating effectively during and after adverse conditions. Ensures continuity of critical operations during cyber incidents. Reduces the impact of attacks by containing and mitigating them swiftly. Enhances the overall security posture by being prepared for various attack scenarios. Note: this is not the same as threat-aware and indeed can be threat agnostic.

Strategic Activities:

- Intrusion Detection and Response: Implement advanced intrusion detection and response mechanisms to mitigate cyber-attacks in real-time.
- Adaptive Defense Mechanisms: Develop adaptive defense mechanisms that can adjust to evolving threats and conditions dynamically.
- Resilient Communication Protocols: Utilize resilient communication protocols to maintain integrity and availability even under attack.

4. Creation of New Metrics and Measures for Resilient Response

Developing new metrics and measures to better understand, monitor, and control resilient responses in real-time. Current metrics are commonly in terms of reliability and robustness and not well suited for resilience. Additionally, metrics in the cyber space are challenging to map to resilience and current risk. New metrics and measures specifically for tracking and driving resilience is necessary to provide quantifiable data to assess and improve resilience practices, enhance situational awareness, and decision-making during cyber incidents, and facilitate continuous improvement and adaptation of resilience strategies. This will need to better address the social and human elements of cyber-physical systems.

Strategic Activities:

- Resilience Metrics: Establish resilience metrics that quantify the ability to anticipate, withstand, recover, and adapt to cyber threats.

- Continuous Monitoring: Implement continuous monitoring systems to track resilience metrics and provide real-time insights into system health and security.
- Incident Reporting: Develop standardized incident reporting frameworks to capture and analyze data on cyber incidents and responses.
- Benchmarking and Improvement: Use resilience metrics to benchmark against industry standards and identify areas for improvement.

Resilience through Communication Network Design and Operation

Resilience in communication network design and operation is vital for ensuring the continuous delivery of critical services and for enabling robust responses to cyber threats. This strategy involves identifying mission-critical dependencies, implementing proactive and adaptive defense mechanisms, applying Zero Trust principles, and integrating advanced communication technologies like 5G into cyber-physical systems. The key elements of the strategy include the following recommended activities:

1. Discovering and Understanding Mission-Critical Cyber Dependencies

Identifying and understanding the cyber dependencies that are critical for mission success is the first step in enhancing network resilience. Provides a clear understanding of what needs to be protected. Helps prioritize resources and efforts toward safeguarding the most critical elements. Enhances situational awareness and preparedness.

Strategic Activities:

- Dependency Mapping: Create a comprehensive map of all mission-critical systems, networks, and interdependencies.
- Tools for Cyber and Physical Risk Assessment: Physical threats and cyber threats are commonly evaluated independently but real-world threats can and do leverage capabilities from both spaces. Tools to enable analysis across both the physical process and security and cyber security.
- Data Flow Analysis: Analyze data flows to understand how information moves through the network and identify critical nodes and communication paths.
- Stakeholder Engagement: Engage with key stakeholders to ensure a thorough understanding of operational requirements and dependencies.

2. Using Proactive and Adaptive Cyber Defense Mechanisms

Implementing defense mechanisms that can proactively detect and respond to threats based on real-time threat intelligence. Enabling this strategy would enhance the ability to detect and respond to threats in real-time, reduce the dwell time of attackers within the network and provide valuable insights into attacker tactics, techniques, and procedures (TTPs).

Strategic Activities:

- Stronger Threat Intelligence Integration with Tools: Incorporate real-time threat intelligence feeds to inform and update defense strategies. Threat intelligence is a valuable source, but critical infrastructure asset owners often lack the resources to fully utilize it. Integrating these data feeds into more automated tools will alleviate the barrier to entry owner/operators and enable them to benefit from current knowledge of ongoing threat campaigns.

- Deception Technologies as a low barrier to entry defense tool: Use deception technologies to create fake assets and environments that can mislead and trap attackers. Asset owners are often hesitant to deploy new security solutions because of the potential or perceived risks. Deception techniques have low risk of impacting operations and have low false positive rates of detecting unauthorized activities.
- Anomaly Detection: Implement systems to detect abnormal behavior that may indicate a potential attack. The physical process side of cyber-physical systems provides an opportunity to ground and constrain abnormalities in cyber data to help with the base rate fallacy challenge of traditional cyber data.
- Automated Response: Develop automated response mechanisms that neutralize threats and adapt to evolving attack patterns. This can cover the range of automated, autonomic, or even autonomous response with human oversight as needed.

3. Limiting Attack Impact with Zero Trust Concepts in OT Environments

Applying Zero Trust principles to Operational Technology (OT) environments to minimize the risk and impact of cyber-attacks. Through the appropriate application of zero trust concepts, it is possible to significantly reduce the attack surface by limiting unauthorized access, contain and minimize the impact of breaches, and provide enhanced visibility and control over the network. If usability is considered in this approach the fears of OT operators can be overcome and enable connectivity, they aspire to create such as access to OT data in business systems while also improving overall security posture.

Strategic Activities:

- Micro-Segmentation: Isolate network segments to limit the spread of attacks.
- Integration of host and network policy and control: Combining host/application sandboxing with network segmentation for full end-to-end data movement mediation and authorization.
- Continuous Monitoring: Continuously monitor all activities within the network to detect and respond to malicious actions.
- Framework across on-prem and cloud assets: Develop approaches and architectures to securely utilize cloud services in OT to prepare for the increasing reliance on cloud applications and infrastructure.

4. Integrating Cutting-edge Communication Technologies such as 5G into Cyber-Physical Systems

Leveraging advanced technologies like 5G/ Next-Gen to enhances cyber-physical systems capabilities and mitigates associated risks. Through advanced wireless utilization the capabilities and performance of cyber-physical systems can be enhanced to provide greater flexibility and scalability for network operations. It can also enable new use cases and applications that were not possible with previous communication technologies like enabling more localized and peer-to-peer communication in support of more distributed and isolated operation.

Strategic Activities:

- Benefits Analysis: Evaluate the potential benefits of 5G, including higher speeds, lower latency, mobile edge computing, and greater device connectivity.
- Risk Assessment: Assess the security risks associated with 5G, such as new attack vectors and increased complexity. This includes the 5G network and/or the devices connected to the network.
- Secure Integration: Develop strategies to securely integrate 5G technology into cyber-physical systems, considering aspects like encryption, authentication, and secure configurations.

- Pilot Programs: Conduct pilot programs to test and refine the integration of 5G in controlled environments before full-scale deployment.
- Collaboration with Providers: Work closely with 5G service providers to ensure that security measures are robust and align with operational requirements.

Resilience through Operator Response Support

Resilience in operator and responder support is essential for maintaining robust defenses and swift responses to cyber threats, particularly in cyber-physical systems. This strategy includes comprehensive training exercises, the development of advanced tools for system emulation and data analysis, and a framework for integrating human decision-making with autonomous control systems, all while promoting collaboration with industry stakeholders. The key elements of the strategy are:

1. Resilience by Training and Preparedness Exercises

Conducting regular training exercises to prepare operators and responders for realistic cyber-attack scenarios. Enhances the readiness of operators and responders to manage actual cyber incidents. Identifies gaps in the current response plans and allows for continuous improvement. Builds confidence and competence in handling cyber threats.

Strategic Activities:

- Simulated Cyber Attacks Exercises: Use high-fidelity simulations and emulations of cyber-physical environments to create realistic attack scenarios. Continuing the investment in programs like DHS CELR or DOE GridEx provides unique and valuable experience to practice and discover the gaps in current resilient response plans. Ongoing support and expansion of these and other like programs is important to make sure the nation's critical infrastructure force is ready to implement resilience measures when necessary.
- Create advanced simulated exercise environment tools to increase the reality of capabilities: In support of the above, continued development of emulation and testbed capabilities for cyber-physical systems is important. This includes developing new models and scenarios, along with the necessary infrastructure and tools, to make training and experimentation environments efficient and high-fidelity.
- Develop TTP test creation pipeline: Threat intelligence continually generates new advisories and CVE reports which are currently not easily operationalized. Develop standards and processes for generating executable TTPs that can be used by industry to validate if their defenses can detect and respond to TTPs detecting in their systems as well as can be used in the training and preparedness exercises to make sure participants are seeing the evolving and most relevant behaviors in the wild.
- National Data Repository of Useful Datasets: The availability of useful datasets has been, and continues to be, a challenge for the research and development industry. Creating national repositories to curate and host high-quality datasets is important to tackle this challenge. National laboratories are the perfect independent broker for such activities to provide the trust to industry to share real datasets but also as a place to generate realistic benchtop/synthetic datasets to fill the repository. Continued support is necessary to address the longitudinal challenge of cyber security where the cyber space and threats evolve rapidly, and aging datasets become less relevant and applicable.

2. Developing Tools for Cyber-Physical System Emulation

Create tools that support efficient and high-fidelity operation of cyber-physical system environment emulation. Provides a safe and controlled environment to test and refine response strategies. Enhances the understanding of how cyber-physical systems react under attack conditions. Supports the development of more resilient system configurations and defense mechanisms.

Strategic Activities:

- Emulation and Experimentation Platforms: Develop advanced emulation platforms that accurately replicate the operational environment of cyber-physical systems. Current capabilities across the national complex can model within specific sectors of critical infrastructure. Expanding and enabling cross-sector and high scale environments are important to enable the research and exercises to optimally prepare for the inevitable future cyber-attacks.
- Scenario Libraries: Build comprehensive libraries of attack scenarios and system responses to test various conditions and threats. Some sectors, such as the power industry, already excel at developing prototypical modeled environments to support research and development. Investments to build out similar capabilities across all the critical infrastructure is important. Also, standardizing modeling languages and protocols will enable sharing off test environments to tackle the reproducibility challenge in research and enable supporting the demand of exercises.
- Integration with Training Programs: Use these emulation tools in training exercises to provide realistic and immersive experiences for operators.

3. Enhancing Analyst Efficiency through AI/ML

Developing tools to make analysts more effective and efficient in parsing, understanding, and responding to cyber-physical data and potential attacks using AI/ML advancements. Increases the speed and accuracy of threat detection and response. Reduces the workload on human analysts by automating routine and complex tasks. Enhances situational awareness and decision-making capabilities.

Strategic Activities:

- AI/ML Integration: Research and development in AI/ML approaches applied to cyber-physical specific use cases. This strategy includes efforts to help guide research around promising and inappropriate applications of AI/ML. There is a lot of misinformation on what AI/ML is capable of and the domains around which it is effective. Providing national guidance documents like those that have been developed for other hot topics like block chain[1] would help reduce the wasted investments and effort. Continued human machine teaming research is needed for specific cyber defense uses.
- Automated Threat Prediction and Detection: Develop AI/ML-driven tools to automatically predict, detect and categorize potential threats and anomalies.
- Analysis Decision Support Systems: Create intelligent decision support systems that provide actionable insights and recommendations to analysts. Use LLM advancements to lower the technical barrier to entry for industry engineers to protect their systems. Create data fusion and blending capabilities to bring together the multiple cyber, physical, and business data to provide a holistic real time system understanding.
- Workforce Development: Cross-train cybersecurity and AI experts to foster collaboration and help bring together a common understanding between the two domain experts.

[1] https://nvlpubs.nist.gov/nistpubs/ir/2018/nist.ir.8202.pdf

- Information Sharing: Create a for sharing information on the safety and security of cyber-physical systems that employ AI technologies. (Along the lines of the National Critical Infrastructure Observatory outlined by the President's Council of Advisors on Science and Technology.[1])

4. Integrating Human Decisions with Autonomous Control Systems

Developing a framework for when and how human decisions appropriately integrate with new distributed and autonomous control systems. Ensures a balanced integration of human intuition and machine precision in decision-making. Enhances the resilience and effectiveness of control systems through coordinated efforts. Provides flexibility to adapt to different types of cyber threats and scenarios.

Strategic Activities:

- Decision-Making Protocols: Establish clear protocols that define the roles and responsibilities of human operators and autonomous systems. These definitions become increasingly important as proactive and adaptive mitigations and distributed control becomes deployed.
- Human-Machine Interfaces: Develop intuitive interfaces for seamless interaction between human operators and automated systems. Current HMI systems rely on the user to interpret the data and determine intent. New human-machine teaming models are required for systems making decisions independent of operators, ensuring the human remains updated and aware of necessary information using new HMI capabilities.
- Adaptive Decision Frameworks: Implement adaptive frameworks where human and machine decision-making processes are dynamically adjusted based on the context and severity of the situation. Expecting a human to take over in extreme edge cases is unreasonable if the human isn't involved throughout the process. Creating the frameworks on when and how to hand over system control is needed.

5. Industry Collaboration and Integration

Ensure that industry is involved and integrated as advisors and collaborators at every step of the national strategy. Leverages industry expertise and innovation to enhance national resilience efforts. Promotes the adoption of best practices and standards across different sectors. Ensures the alignment of national resilience strategies with real-world needs and challenges. Private industry does not bear responsibility for national resilience and security, but they are currently bearing the costs. We need to move beyond talking and into the realm of action and joint developments.

Strategic Activities:

- Public-Private Partnerships: Foster collaborations between government agencies, private sector entities, and academic institutions.
- Industry Advisory Boards: Create advisory boards composed of industry experts to provide insights and guidance on resilience strategies.
- Joint Research Initiatives: Support joint research and development initiatives to advance cybersecurity practices and technologies.

[1] President's Council of Advisors on Science and Technology, Strategy for Cyber-Physical Resilience: Fortifying Our Critical Infrastructure for a Digital World (Washington, DC: Executive Office of the President, 2024), www.whitehouse.gov/wp-content/uploads/2024/02/PCAST_Cyber-Physical-Resilience-Report_Feb2024.pdf.

**Pacific Northwest
National Laboratory**

www.pnnl.gov

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Parv Venkitasubramaniam

Jeff Heflin

Paolo Bocchini

Jessecae Marsh

Ted Ralphs

Rick Blum

Brian Davison

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Response to Networking and Information Technology Research and Development Request for Information on a National Plan for Cyber-Physical Systems Resilience (89 FR 78915)

Author Names: Parv Venkitasubramaniam, Jeff Heflin, Paolo Bocchini, Jessecae Marsh, Ted Ralphs, Rick Blum, Brian Davison.
Institution: Lehigh University
Statement: This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience RD Strategic Plan and associated documents without attribution

Explainable AI for Disaster Response

Despite vast amounts of data available, real-time data-driven decision making during extreme events such as natural disasters or large-scale cyber-attacks continues to be challenging, for multiple reasons. First, extreme events are infrequent and each event is unique. Available data overwhelmingly pertain to normal conditions, or at best to other real or simulated events; computational models trained using such data are not reliable under disaster conditions that are previously unseen by the models [8, 12]. Second, as a disaster unfolds, the state of infrastructure and underlying user behavior evolve rapidly [11]. Decision-making tools need to continually adapt to changing physical conditions and data patterns, without requiring high real-time computation. Third, emergency response actions during disasters are taken by human operators who bear professional liability for their decisions [6]. Any data-driven tool and outcome should be explainable to the operators, to gain their trust. Without such tools, emergency responders during a disaster often rely on experience or on pre-engineered solutions with limited usage of real-time data. Faster emergency response can not only lower the human costs of disasters [1, 10] but also the financial costs [2]. Addressing the aforementioned challenges requires fundamental research advances that cross-cuts multiple disciplines that include infrastructure modeling and disaster management, data science and machine learning, knowledge representation, efficient and explainable computation and cognitive psychology.

In this regard, there are four major areas of research that we propose:

Reliable prediction in unseen and unpredictable conditions: A major challenge in ML, referred to as out-of-distribution generalization, is to develop models that provide reliable outcomes when test data patterns are previously unseen during the training phase. Despite various approaches designed for ML in research areas such as vision [4, 13, 15] and text [14], addressing the challenge in dynamical and cyber-physical systems remains open. Most existing models for time-series prediction assume data are independent and identically distributed [3,5,9]. Newer machine learning approaches and computational architectures that adhere to underlying physical laws are required to improve prediction performance under conditions significantly different from those encountered in training data. Approaches that can deal with rapidly changing conditions and missing or false data as is common under extreme scenarios are crucial to enable robust and reliable prediction under these scenarios.

Adaptive Computation under rapidly evolving conditions: Computational algorithms for decision making in cyber-physical systems are typically designed based on static or predictable system states, and are not trained to include information such as impending risks of damage from an evolving disaster or a cyberattack. Modern catastrophe modeling approaches offer the opportunity to assess in near-real-time the probability of loss of functionality of system components and processes, and these measures along with the predicted system states must be used in performing the needed computations and optimized decision-making in an evolving disaster. Fast optimization and adaptive computational methods need to be developed to apply in disaster scenarios which can quickly adapt to changing conditions of the system and environment.

Explanations for computational outcomes: From a cognitive psychological perspective, very little is known about what kind of explanations for data-driven models can gain trust of actual users of the methods and outcomes. Explainable AI methodologies that have been developed, particularly for machine learning solutions, are focused on “simplifying” the computational process or outcomes rather than considering a holistic view of the user in the context of the application [7]. Different kinds of explanations have been proposed for computational models such as feature-importance, model interpretation, example-based, etc., are limited to generic ML models, and not tailored to domain-specific conditions and constraints, and additionally, the approaches are not adaptable to changing system or environmental conditions. Rather than a computation-only approach, what is required is a holistic framework to derive explanations that are computationally accurate, contextually interpretable, and guided by a use-inspired framework of human perception and decision-making. This requires bringing together concepts and tools from knowledge representation, cognitive psychology, data science, optimization and machine learning, to build explainable representations of domain knowledge, derive rules to contextualize the outcomes of data-driven decisions, and use cognitive psychological studies to formally understand the perception and use of explanations during extreme events to help formulate explanations that will be adopted and used.

Human-in-the-loop Testing Platform Given the challenges with real-time disaster testing of developed methods, there is a need to build a human-in-the-loop experimental platform that can simulate disaster or cyberattack scenarios to generate data on the real-time impact on infrastructural systems, as a function of the unfolding disaster as well as the human decision making in its aftermath. Such an experimental platform should have the ability to provide risk analysis on the evolving disaster scenario and the time-varying state of the infrastructure. The testing platform should provide the necessary feedback to validate the methodological outcomes as well as the validity and usability of the explanations by human operators.

References

[1] Saman Ahmadi, Guido Tack, Daniel D Harabor, and Philip Kilby. A fast exact algorithm for the resource constrained shortest path problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12217–12224, 2021.

[2] Daniel Brent and Louis-Philippe Beland. Traffic congestion, transportation policies, and the performance of first responders. *Journal of Environmental Economics and Management*, 103:102339, 2020.

[3] Mark Granroth-Wilding and Stephen Clark. What happens next? event prediction using a compositional neural network model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[4] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8349, 2021.

[5] Balázs Hidasi and Alexandros Karatzoglou. Recurrent neural networks with top-k gains for session-based recommendations. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 843–852, 2018.

[6] Nan D Hunter. *The law of emergencies: public health and disaster management.* Butterworth-Heinemann, 2017.

[7] Uday Kamath and John Liu. *Explainable artificial intelligence: An introduction to interpretable machine learning.* Springer, 2021.

[8] Vaishnavh Nagarajan, Anders Andreassen, and Behnam Neyshabur. Understanding the failure modes of out-of-distribution generalization. In *International Conference on Learning Representations*, 2020.

[9] Zhi Qiao, Shiwan Zhao, Cao Xiao, Xiang Li, Yong Qin, and Fei Wang. Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction. In *Proceedings of the twenty-seventh international joint conference on artificial intelligence*, 2018.

[10] Rocío Sánchez-Mangas, Antonio García-Ferrrer, Aranzazu De Juan, and Antonio Martín Arroyo. The probability of death in road traffic accidents. how important is a quick medical response? *Accident Analysis & Prevention*, 42(4):1048–1056, 2010.

[11] Wenjuan Sun, Paolo Bocchini, and Brian D Davison. Applications of artificial intelligence for disaster management. *Natural Hazards*, 103(3):2631–2689, 2020.

[12] Yue Sun, Chao Chen, Yuesheng Xu, Sihong Xie, Rick S Blum, and Parv Venkitasubramaniam. On the generalization discrepancy of spatiotemporal dynamics-informed graph convolutional networks. *Frontiers in Mechanical Engineering*, 10:1397131, 2024.

[13] Florian Wenzel, Andrea Dittadi, Peter Gehler, Carl-Johann Simon-Gabriel, Max Horn, Dominik Zietlow, David Kernert, Chris Russell, Thomas Brox, Bernt Schiele, et al. Assaying out-of-distribution generalization in transfer learning. *Advances in Neural Information Processing Systems*, 35:7181–7198, 2022.

[14] Linyi Yang, Yaoxiao Song, Xuan Ren, Chenyang Lyu, Yidong Wang, Lingqiao Liu, Jindong Wang, Jennifer Foster, and Yue Zhang. Out-of-distribution generalization in text classification: Past, present, and future, 2023.

[15] Chongzhi Zhang, Mingyuan Zhang, Shanghang Zhang, Daisheng Jin, Qiang Zhou, Zhongang Cai, Haiyu Zhao, Xianglong Liu, and Ziwei Liu. Delving deep into the generalization of vision transformers under distribution shifts. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 7277–7286, 2022.

Federal Register Notice: 89 FR 78915, Federal Register :: Networking and Information Technology Research and Development Request for Information on Cyber-Physical Systems Resilience Research, September 26, 2024.

Request for Information on the National Cyber-Physical Systems Resilience Plan

Vinayak S. Prabhu

Indrajit Ray

Indrakshi Ray

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Formal Specifications for Resilient CPS Design

Vinayak S. Prabhu
Colorado State University
[redacted]

Indrajit Ray
Colorado State University
[redacted]

Indrakshi Ray
Colorado State University
[redacted]

October 26, 2024

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Cyber-Physical Systems Resilience R&D Strategic Plan and associated documents without attribution.

We argue for formal methods based frameworks to facilitate the development of resilient Cyber-Physical Systems (CPS). One of the core elements of formal methods is the rigorous specification of system requirements. Such requirements are specified in frameworks such as timed and untimed temporal logics, automata based frameworks, or process calculii frameworks. In the context of CPS, such frameworks have been augmented to incorporate signals which are functions specifying the evolution of signal values over time. Typically the correctness of systems is guaranteed under certain environment hypotheses — the system is only guaranteed to be correct provided the environment is not *too* antagonistic. Both the environment hypotheses and the system guarantees are given in the frameworks mentioned above in a reactive setting where the system is continually interacting with the environment.

For resilient CPS design [HSK19, Fir19], we need to reason about degraded modes of operation, and we argue the specification frameworks need to be explicit concerning the degraded modes. For example, a component guarantee could be that under environment hypotheses H_1, it can guarantee behavior G_1, and under a degraded mode H_2 where the environment is more hostile (for instance, under increased network latency due to the system being under attack), it can guarantee behavior G_2.

For a modular CPS topological setting where the CPS is a composition of several modules and one/many of the modules may be under attack leading to degraded modes of operation, the research problems include what frameworks need to be developed to specify the hypotheses and the guarantees to facilitate rigorous reasoning about the impact chain of degradation of one component on the other components in the system. If the system were monolithic, one could have the hypotheses in one framework, for instance as parameters such as network delay, the CPU speed of the underlying execution platform, the actuator delay etc, and the system guarantees could be in standard formal methods based frameworks. This would allow one to specify system behaviors under various forms of attacks affecting network delays, CPU throttling due to temperature attacks, actuator attacks etc. However, such a coarse formalism is not amenable towards analysis of how degradation of one component affects other components.

The hypotheses, as well as the guarantees ideally need to be connected to quality of service (QoS) measures of the system. A purely Boolean logical framework (where a property is either satisfied or not) might not be appropriate in many settings; a more desirable quantitative framework would indicate how well a property is satisfied, with the "how well" number being an interpretable QoS measure.

The explicit specification of such hypotheses, guarantees, and a formal framework for analyzing the network effects of component degradation would highlight which are the most attack vulnerable components that need to be fortified the most, or even that the CPS component topology needs to be changed. For example, due to the reactive nature of CPS where components are continually interacting with one another, the initial degradation of component X's operation may get amplified over time due to component X negatively affecting component Y, and the resulting degradation in component Y further degrading X over time, in a feedback loop.

The above example also illustrates the need for reasoning about dynamically changing the topology of software functionality amongst the CPS modules using computational models such as Bigraphs [Mil09].

A design of resilient systems involves 3 Rs: *Recognition*, *Resistance*, and *Recovery.* The property specification frameworks need to be expressible enough to be dynamically expressive in the following manners:

1. Express a monitoring framework that recognizes when a module is under attack.

2. Express the desired property that the effect of quantified short term disruptions lasts for only a quantified bounded time interval, and quantify these effects.

3. Express the transient properties that modules must satisfy in the period when the system recovers or adjusts (eg., when the environment hypothesis switches back from a degraded mode H_2 back to H_1).

4. Express the stabilization properties which specify the system adjusting to various hypotheses changes.

The system architecture also needs to concretize the mechanism for *restoration* – for instance the mechanism for restoring the original system state by restoring/replacing any particular modules that were compromised. There needs to be a rigorous specification and checking of the transient behavior of the system when the compromised modules are being restored/replaced.

A more resilient system provides more functionality than a non-resilient one, and thus the software implementing such a resilient system would necessarily be larger. However, more complex software is more vulnerable to attacks, which in turn makes the system *less* resilient. Thus the desired functionality increase of resilient systems must be implemented with care. For instance, the various modes of component operations under various levels of environment degradation could mathematically correspond to restricted models of computation, which are easier to implement in software, and also easier to formally verify as compared to general software. Such model restrictions, even though they might be inferior to general computational models for performance guarantees, might be preferable as they reduce the attack surface itself. Thus, the model restriction – model performance guarantee trade-off for CPS is also an important consideration for resilient CPS design.

Lastly, a resilient CPS system design needs to account for *human factors*: does the system alert human operators at several levels with sufficient redundancies when modules are suspected of being compromised; are the alerts descriptive and understandable; and is the system robust when multiple human operators attempt to intervene in contradictory manners.

References

[Fir19] Donald Firesmith. System resilience: What exactly is it? Carnegie Mellon University, Software Engineering Institute's Insights (blog), Nov 2019. Accessed: 2024-Oct-26.

[HSK19] Md Ariful Haque, Sachin Shetty, and Bheshaj Krishnappa. Cyber-physical system resilience. In *Complexity Challenges in Cyber Physical Systems*, chapter 12, pages 301–337. John Wiley & Sons, Ltd, 2019.

[Mil09] Robin Milner. *The Space and Motion of Communicating Agents*. Cambridge University Press, 2009.