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

“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.”

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.