POSTER: A Common Framework for Resilient and Safe Cyber-Physical System Design

Luyao Niu
University of Washington
Seattle, WA, USA
luyaoniu@uw.edu

Abdullah Al Maruf
University of Washington
Seattle, WA, USA
maruf3e@uw.edu

Andrew Clark
Washington University in St. Louis
St. Louis, MO, USA
andrewclark@wustl.edu

J. Sukarno Mertoguno
Georgia Institute of Technology
Atlanta, GA, USA
karno@gatech.edu

Radha Poovendran
University of Washington
Seattle, WA, USA
rp3@uw.edu

ABSTRACT

Cyber-physical systems (CPS), which are often required to satisfy critical properties such as safety, have been shown to be vulnerable to exploits originating from cyber and/or physical sides. Recently, novel resilient architectures, which equip CPS with capabilities of recovering to normal operations, have been developed to guarantee the safety of CPS under cyber attacks. These resilient architectures utilize distinct mechanisms involving different parameters and are seemingly unrelated. Currently, the analysis and design methods of one novel resilient architecture for CPS are not readily applicable to one another. Consequently, evaluating the appropriateness and effectiveness of a set of candidate resilient architectures to a given CPS is currently impractical. In this poster, we report our progress on the development of a common framework for analyzing the safety and assessing recovery performance of two or more resilient architectures intended for CPS under attacks. We formulate a hybrid model as a common representation of resilient architectures. Our insight is that the resilient architectures have a shared set of discrete states, including vulnerable, under attack, unsafe, and recovery modes, which can be mapped to the discrete states of the unifying hybrid model. The hybrid model enables a unified safety analysis. We parameterize the required behaviors for the cyber and physical components in order to guarantee safety. The parameters then inform the development of metrics to measure the resilience of CPS. For CPS consisting of multiple heterogeneous components, we show that the effect of interconnections on the spatial and temporal parameters can be quantified efficiently, allowing a compositional approach to the safety verification of large-scale CPS.

CCS CONCEPTS

• **Computer systems organization → Embedded and cyber-physical systems; Dependable and fault-tolerant systems and networks**.

ASIA CCS '23, July 10–14, 2023, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0098-9/23/07.
https://doi.org/10.1145/3579856.3592826

KEYWORDS

Cyber-physical systems, adversary, resilient architectures, safety

ACM Reference Format:
Luyao Niu, Abdullah Al Maruf, Andrew Clark, J. Sukarno Mertoguno, and Radha Poovendran. 2023. POSTER: A Common Framework for Resilient and Safe Cyber-Physical System Design. In *ACM ASIA Conference on Computer and Communications Security (ASIA CCS '23), July 10–14, 2023, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3579856.3592826

1 INTRODUCTION

CPS consist of cyber components (e.g., communication and computation) that are coupled with physical components (e.g., sensors and actuators). CPS have been widely deployed, including in critical infrastructures [10] and automobiles [5]. The couplings between the cyber and physical components make CPS vulnerable to attacks. Such attacks have been reported across multiple real-world applications such as automobiles and power systems [5, 10]. The operation of CPS in adversarial environments can affect the satisfaction of required performance objectives such as safety, causing damage to the system or harm to human operators [5, 10]. Moreover, adversaries can continuously adapt their behaviors to successfully bypass existing fault-tolerant mechanisms such as Simplex [9] by exploiting software vulnerabilities and developing new attacks.

Recent research has focused on the development of a family of resilient architectures [1–4, 6–8] to provide guarantees of desired performance when CPS are subject to attacks. These architectures leverage an insight that an adversary will not be able to persistently compromise the CPS if cyber components of the system are restored to a '*clean*' state. The status of cyber component changes at a higher frequency compared to the state evolution of the physical component whose dynamics are governed by the laws of physics. Therefore, the resilient architectures enable CPS to tolerate temporary loss of or corrupted control inputs caused by cyber attacks without causing violations of desired performance.

Although every individual resilient architecture is promising on its own, they each have distinct design parameters that will need to be configured. For instance, [2] restarts the cyber component with a tunable frequency regardless of the input being compromised or not, whereas [6] restores the system once an erroneous input is received. Consequently, practitioners may lack guidelines on strategies to tune parameters of each architecture when applying

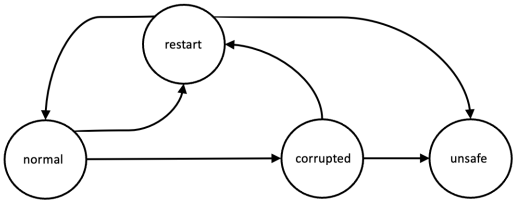

Figure 1: Our proposed hybrid model to abstract resilient architecture YOLO. The circles are discrete states of the hybrid model, which comprise the cyber statuses and an absorbing state $unsafe$ modeling safety violations. The arrows represent the transitions among the discrete states.

them to heterogeneous CPS operated in distinct environments. At present, there is no principled methodology to quantify and evaluate the effectiveness and appropriateness of the resilient architectures.

In this poster, we report our progress on the development of a new common framework that will enable the simultaneous analysis and design of multiple resilient architectures. Our critical insight is that at each time instant, we can use the availability of the cyber component to categorize the cyber status of CPS employing any resilient architecture. We combine the discrete transitions along with the continuous dynamics of the physical component to develop a hybrid model as a representation of the resilient architectures. The proposed hybrid model informs a unified parameterized method to verify the safety of CPS employing distinct resilient architectures.

Our research has a three-pronged focus. We develop a hybrid model as a common representation for the resilient architectures in Section 2. Section 3 shows how the hybrid model allows a unified analysis for the resilient architectures. We present how the hybrid model and unified analysis can be extended to large-scale interconnected CPS in Section 4. Section 5 concludes this paper.

2 HYBRID MODEL REPRESENTATION OF RESILIENT ARCHITECTURES

We consider a CPS consisting of a cyber and physical component. The CPS is subject to a safety constraint, requiring the physical component to remain in a safety set $C = \{x : h(x) \geq 0\}$, where h is continuously differentiable and x is the state of the physical component. The states of physical components of the CPS change with time, and are additionally influenced by an input signal provided by the cyber component. The cyber component is subject to cyber attacks launched by an intelligent adversary The adversary can exploit the cyber vulnerabilities and intrude into the cyber component. Upon a successful intrusion, it gains root access to the system. As a consequence, the adversary corrupts the control input issued by the cyber component to disrupt the operation of CPS and cause safety violations. Throughout this paper, we assume that there exists a trusted and uncompromised image, from which the cyber component could be restored to the attack-free state.

We illustrate how a resilient architecture is abstracted using a hybrid model. We use a resilient architecture named YOLO (You Only Live Once) [2] as an example. We then present our proposed hybrid model as a common representation for the resilient architectures.

YOLO periodically restarts the cyber component to prevent an adversary from persistently compromising the input issued to the

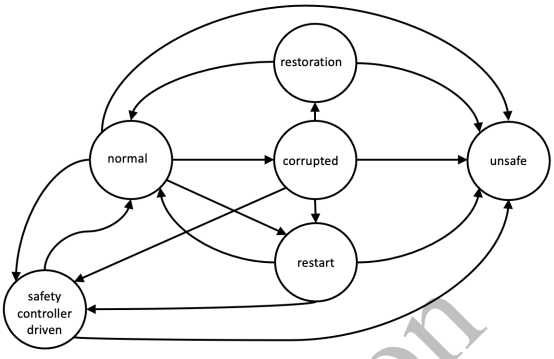

Figure 2: Our proposed hybrid model as a common representation of the resilient architectures including [1–4, 6–8]. The circles are discrete states. The arrows represent the transitions among states.

physical component. Therefore, if the adversary has not successfully intruded into the system, then the cyber component switches between two statuses in $\{normal,\ restart\}$. At the *normal* status, the cyber component is attack-free, and issues desired control input to the physical component. Status *restart* models the phase when the cyber component reboots to reset to a 'clean' state. When the cyber component is at *restart* status, it issues no control input to the physical component. If the adversary has intruded into the system, then the cyber component switches among *normal*, *corrupted*, and *restart* statuses. Status *corrupted* models the case where the cyber component issues compromised control input. To summarize, YOLO defines a set of periodic transitions among three cyber statuses $\{normal, restart, corrupted\}$. Note that the states of physical component evolve continuously over time based on the control input issued by the cyber component. We can thus formulate any CPS employing YOLO as the hybrid model shown in Fig. 1, where $unsafe$ models the scenarios where safety constraint is violated.

The key insight informing the design of our hybrid model is that the resilient architectures share a finite set of cyber statuses, i.e., $\{normal, corrupted, restart, restoration, safety\ controller\ driven\}$. Here, status *restoration* models the scenario where the cyber component is restored by using some redundant backup copy, instead of directly rebooting. Status $safety\ controller\ driven$ models CPS equipped with a safety controller that is guaranteed to be attack-free, and can be invoked when the cyber component is compromised by the adversary. Leveraging this insight, we construct a hybrid model shown in Fig. 2 as a common representation of the family of resilient architectures. The cyber statuses combined with state $unsafe$ comprise the discrete state set of the hybrid model. The resilient architecture employed by the CPS determines the transitions among the states, which are depicted as arrows in Fig. 2.

3 A UNIFIED ANALYSIS OF SAFETY

This section describes how the proposed hybrid model in Fig. 2 enables a unified safety analysis of CPS employing any resilient architectures. Given the hybrid model as shown in Fig. 2, we can convert the problem of guaranteeing safety of CPS under cyber attacks to an equivalent problem of ensuring the hybrid model

never reaches state $unsafe$. However, solving the converted problem is still computationally intensive, especially when the physical component is governed by high-dimensional nonlinear dynamics.

We propose a decompositional approach, which constrains the physical component to remain within a subset of the safety set for each cyber status. Given a status, we find its corresponding subset by quantifying (i) the maximum variation of the level set $h(x)$ at the status, and (ii) the time duration for which the hybrid model remains at the status. For any sequence of transitions on the hybrid model, we then sum up the maximum variations of all traversed statuses, and verify whether the physical component is safe or not.

We illustrate our approach using an example on the altitude control of a drone, as shown in Fig. 3. Suppose that the drone utilizes YOLO as the resilient architecture to mitigate cyber attacks. The drone is given a safety constraint, requiring the altitude of the drone with respect to ground to be nonnegative to avoid crash. We consider a sequence of cyber statuses $corrupted, restart, normal, \ldots$ as shown along the 'Time' axis in Fig. 3. The heights of the yellow, cyan, and purple regions in Fig. 3 represent the maximum variations for statuses $corrupted$, $restart$, and $normal$, respectively. As shown in Fig. 3, by requiring the drone to stay within the level sets for each cyber statuses, we can guarantee that the drone never crashes (shown by the solid curve) in the presence of cyber attacks.

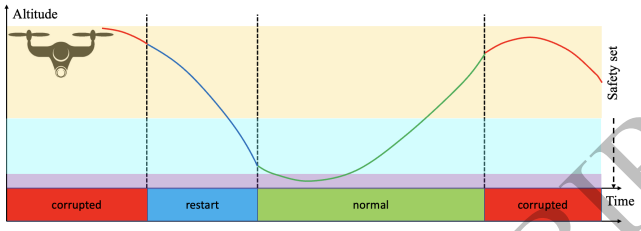

Figure 3: The altitude control of a drone implementing YOLO [2]. The drone needs to maintain nonnegative altitude to avoid crashing. The 'Time' axis shows a sequence of cyber statuses $corrupted, restart, normal, \ldots$. For each status, we let the physical component remain within a subset of the safety set. At statuses $corrupted$, $restart$, and $normal$, we ensure safety by requiring the drone not to fall below the lower boundary of the yellow, cyan, and purple regions, respectively.

4 COMPOSITIONAL RESILIENCE INDEX FOR SAFETY OF INTERCONNECTED CPS

Many real-world applications such as formation of drones and power systems consist of multiple heterogeneous CPS that are connected with each other. In this setting, each CPS may have different dynamics of their physical components, and employ distinct resilient architectures to mitigate cyber attacks, making safety verification computationally challenging. In this section, we show that our hybrid model and safety analysis can be composed for interconnected CPS, enabling efficient safety verification.

We observe that Section 3 characterizes CPS adopting resilient architectures in both spatial and temporal domains (see Fig. 3 for an example). We can thus use the corresponding spatial and temporal parameters, termed *resilience index*, to specify the required behavior of the CPS to guarantee safety. The developed resilience index allows one to compare and evaluate the effectiveness of resilient architectures. Given the interconnections among the CPS and the current values of their resilience indices, the effect of interconnections on the resilience index of each CPS can then be quantified by solving a system of linear inequalities. Thus, the problem of verifying safety for large-scale and societal-scale interconnected CPS is translated to checking the feasibility of linear inequalities, which can be efficiently completed using commercial solvers.

5 CONCLUSION

In this paper, we reported our progress on developing a hybrid model as a common representation of the family of resilient architectures. We proposed a unified method for safety analysis for CPS employing the resilient architectures. We developed a metric called resilience index to evaluate the effectiveness of resilient architectures. The resilience index enabled a compositional approach to safety verification of large-scale CPS. For future directions, we will study the problem of using our hybrid model to guarantee other properties such as stability. We will further investigate the synthesis of new resilient architectures by using the developed hybrid model and unified analysis.

ACKNOWLEDGMENTS

We are grateful to Prof. Bhaskar Ramasubramanian from Western Washington University for the discussions. This work was supported by the AFOSR grants FA9550-20-1-0074 and FA9550-22-1-0054, by the Office of Naval Research grant N00014-20-1-2636, and by the BIRD Foundation: Israel-US Energy Center, Cyber Topic.

REFERENCES

[1] Fardin Abdi, Chien-Ying Chen, Monowar Hasan, Songran Liu, Sibin Mohan, and Marco Caccamo. 2018. Guaranteed physical security with restart-based design for cyber-physical systems. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. ACM/IEEE, 10–21.

[2] Miguel A Arroyo, M Tarek Ibn Ziad, Hidenori Kobayashi, Junfeng Yang, and Simha Sethumadhavan. 2019. YOLO: frequently resetting cyber-physical systems for security. In *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2019*, Vol. 11009. International Society for Optics and Photonics, 110090P.

[3] Stanley Bak, Deepti K Chivukula, Olugbemiga Adekunle, Mu Sun, Marco Caccamo, and Lui Sha. 2009. The system-level Simplex architecture for improved real-time embedded system safety. In *15th IEEE Real-Time and Embedded Technology and Applications Symposium*. IEEE, 99–107.

[4] Marco A Gamarra, Sachin Shetty, Oscar R Gonzalez, Laurent Njilla, Marcus Pendleton, and Charles Kamhoua. 2019. Dual redundant cyber-attack tolerant control systems strategy for cyber-physical systems. In *IEEE International Conference on Communications*. IEEE, 1–7.

[5] Andy Greenberg. 2015. Hackers remotely kill a Jeep on the highway–with me in it. https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

[6] J Sukarno Mertoguno, Ryan M Craven, Matthew S Mickelson, and David P Koller. 2019. A physics-based strategy for cyber resilience of CPS. In *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2019*, Vol. 11009. International Society for Optics and Photonics, 110090E.

[7] Luyao Niu, Dinuka Sahabandu, Andrew Clark, and Radha Poovendran. 2022. Verifying safety for resilient cyber-physical systems via reactive software restart. In *ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 104–115.

[8] Raffaele Romagnoli, Bruce H Krogh, and Bruno Sinopoli. 2019. Design of software rejuvenation for cps security using invariant sets. In *American Control Conference (ACC)*. IEEE, 3740–3745.

[9] Lui Sha. 2001. Using simplicity to control complexity. *IEEE Software* 18, 4 (2001), 20–28.

[10] Julia E Sullivan and Dmitriy Kamensky. 2017. How cyber-attacks in Ukraine show the vulnerability of the US power grid. *The Electricity Journal* 30, 3 (2017), 30–35.