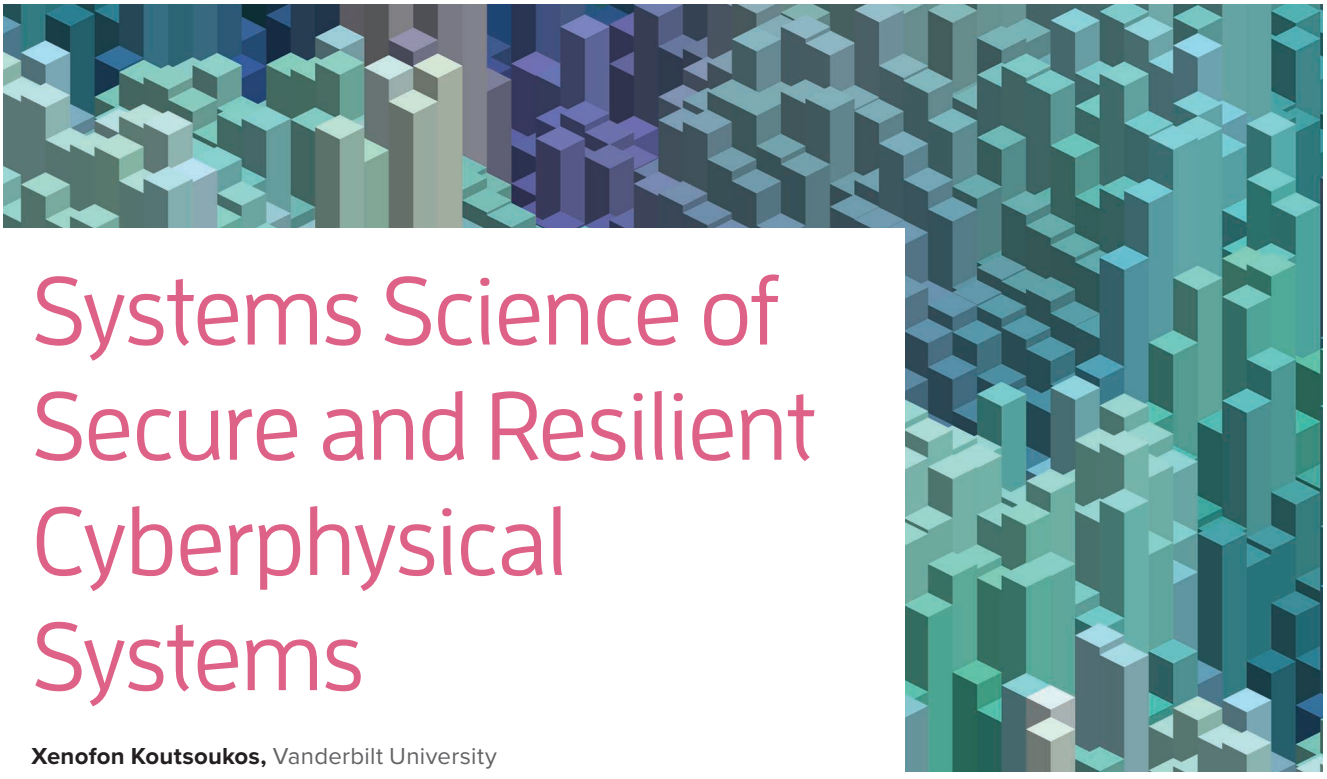
EDITOR **DIMITRIOS SERPANOS**
ISI/ATHENA and University of Patras;
serpanos@computer.org



Systems Science of Secure and Resilient Cyberphysical Systems

Xenofon Koutsoukos, Vanderbilt University

We are in the midst of a pervasive, profound shift in the way humans engineer physical systems and manage their physical environment using networking and IT. Because of these disruptive changes, physical systems can now be attacked through cyberspace, and cyberspace can be attacked through physical means.

According to one of the widely accepted definitions, cyberphysical systems (CPSs) are engineered systems where functionality emerges from the networked interaction of computational and physical processes. The tight integration creates novel systems with revolutionary impacts. This is evident in autonomous vehicles, military platforms, intelligent buildings, smart energy systems, robots, and smart medical devices. Emerging industrial platforms such as the Internet of Things (IoT) are triggering a gold rush toward new markets and creating societal-scale systems that, in addition to the synergy of computational and physical components, interact closely with humans (H-CPSs).

A profound revolution driven by technology and market forces is turning whole industrial sectors into producers of CPSs. This innovation is not about adding computing and communication equipment to conventional products where both sides maintain separate identities. It is about merging computing and networking with physical systems to create new capabilities and product qualities. Whether we recognize it or not, we are at the center of this overwhelming change.

Complex H-CPSs abound in modern society, and it is not surprising that they are a target for attacks. High-profile attacks have been reported in a broad range of systems. For example, researchers have demonstrated the ability to compromise modern automobiles with cyberattacks that can lead to catastrophic physical consequences.[1] Even

Digital Object Identifier 10.1109/MC.2020.2966109
Date of current version: 12 March 2020

under normal conditions, CPSs face complex issues crosscutting many disciplines. Adding cyberattacks in all their insidious variety creates a massive challenge that cannot be neglected due to the potential consequences.

H-CPSs can be organized into abstraction layers dictated by the heterogeneity of their component technologies. Figure 1 shows a simplified view of abstraction layers with distinct architectures, design patterns, composition principles, and vulnerabilities. The physical layer embodies physical components and their interactions. The two cyberplatform layers, network and service platforms, comprise the digital hardware side and include the networks and computation platforms that interact with the physical components through sensors and actuators. The application layer makes up the software components that provide the desired functionality and interfaces. Humans are directly involved in system operation and affect its properties, for example, as operators of vehicles and consumers of services or as adversaries that seek to compromise the system and cause disruption.

The abstraction layers are associated with technologies that have specific vulnerabilities and attack vectors. A critical implication of such complex layered architecture is that the attack surface is very large and heterogeneous. Attacks can include physical destruction, network spoofing, malware, data corruption, and others. Further, attacks are not isolated to a single layer because cross-layer interactions propagate their impact. Attacks on the physical layer can cause anomalies on cyber layers. For example, blocking cooling mechanisms may lead to circuits overheating that, in turn, may cause a shutdown of services implemented by impacted processors. Similarly, cyberattacks can have a physical impact. For example, integrity attacks on sensors may cause incorrect actuation, leading to catastrophic system failures.

CPS security and resilience have attracted considerable attention. Because of their heterogeneity and complexity, existing methodologies are very diverse with different objectives, specifications, and constraints, resulting in a broad body of knowledge.[2] Scientific methods are being used in research efforts to shape technology, practice, and policy in protecting systems from attackers, detecting intrusions, and recovering from compromises. However, scientific methods remain underutilized, and they do not adequately address the interdisciplinary sociotechnical aspects.

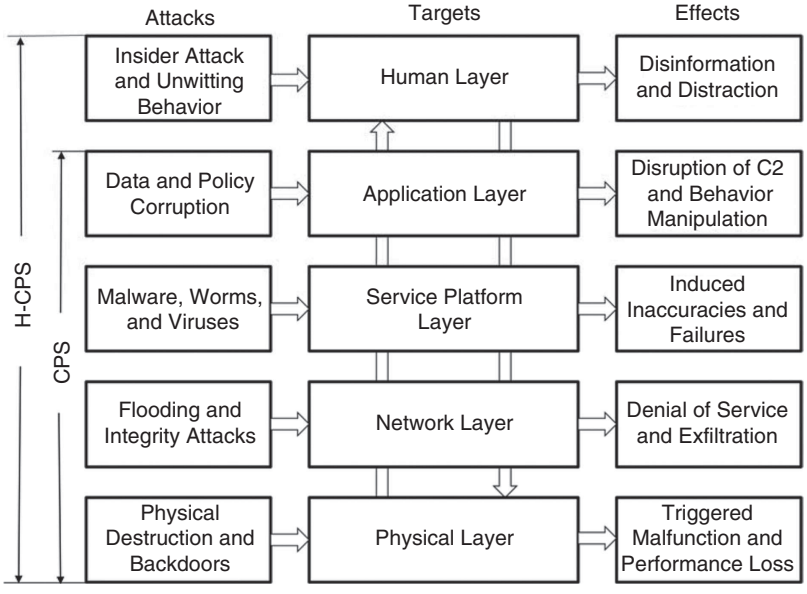


FIGURE 1. The H-CPS abstraction layers.

Beyond the complex structure and interactions, security and resilience properties emerge from complex interrelationships between engineered systems and humans; they are not explained by understanding the individual elements of the system and are highly dynamic in response to changing environments and circumstances. A systems science of secure and resilient CPSs is needed that brings together interdisciplinary research with the goal of identifying, exploring, and understanding patterns of complexity across disciplines and application domains.

DEVELOPING SECURE AND RESILIENT CPSs

Securing CPSs requires developing the principles for security and resilience and using them for system design and management. Methods and tools based on system and game theory, formal methods, data science, incentive engineering, social science, and psychology must be combined to develop integrated solutions that increase our understanding of complex interrelationships, anticipate future conditions, and support decision and policy making. In particular, a systems science seeks intellectual advances in which underlying theories are integrated and abstracted to develop explanatory models. Such explanatory models derived from the underlying theoretical foundations lead to testable hypotheses. Hypotheses are tested using analysis tools as well as simulation and experimentation testbeds to gain a greater understanding of attacks and defenses. Based on collected evidence supporting or falsifying the hypotheses, new insights are obtained, allowing the explanatory models to be refined or updated and, in turn, used for formulating and analyzing new hypotheses.

Such an approach involves significant challenges that are compounded due to significant semantic gaps between 1) the scientific methods used in different disciplines (engineering, policy, economics, and psychology) that are needed to investigate the hard problems

and 2) the different models and representations across abstraction layers and application domains. These challenges call for a systems science that seeks answers to fundamental questions, including the following:

- What are the analytical approaches to model and compose system elements, policies, humans, and adversaries at different layers?
- What are the common semantic domains in which the cross-layer interactions can be described, constrained, and used to compose global security and resilience properties?
- What are the theoretical foundations to analyze the dynamics of CPSs that evolve based on policies, constraints, interactions, and malicious tactical and strategic adversaries?
- What is the theoretical framework that relates the security metrics of local models to those of the global system for different operating conditions?

Transforming the security and resilience of CPSs from a high-risk management practice into an engineering discipline based on systems science is a significant challenge requiring a collaborative and integrative effort. The potential impact is broad and includes

- developing a systematic body of knowledge with both strong theoretical and empirical underpinnings using simulation and experimental testbeds to drive the engineering of secure and resilient CPSs that can resist cyberattacks
- an increased understanding of complex system design and integration, which is the key enabler for building a secure and resilient CPS
- establishing the principles for scalable secure computing and network platforms for CPSs as well as automated methods for reasoning about their security properties.

FOUNDATIONS OF CPS SECURITY AND RESILIENCE

The basic components of information security are confidentiality, integrity, and availability, and these have been used extensively to shape the science and technology of computer security. What are the main components of CPS security and resilience? How can we shape research efforts in developing secure and resilient architectures? How do we organize, analyze, integrate, and evaluate the broad range of available techniques?

Security and resilience can be achieved by either passive or active methods. *Passive methods* aim to establish properties that are inherently robust against classes of uncertainties and secure against cyberattacks. Examples include decreasing safety and security risks by increasing safety margins, hardening access control policies, or using longer encryption keys. *Active methods* refer to the ability of a system to respond to attacks that imply some form of reflexive or deliberative control. Reflexive methods employ a monitor-response scheme, while deliberative methods expand anomaly detection with root cause analysis, isolation, recovery planning, and mitigation actions using detailed information about the structure and expected behavior of the system. A systems science aims at analyzing, designing, and integrating passive and active methods for improving the security and resilience of CPSs.

This problem is particularly hard because cyberattacks may be coordinated on different layers and combined with physical attacks to achieve maximum damage. A necessary but challenging step is to develop a framework for modeling cyber- and cyberphysical attacks at multiple abstraction layers. Attack models integrated with system models can be used to develop methods to improve security and resilience by determining investments in various security mechanisms. To capture both the physical and cyber aspects, the framework must build on multiple disciplines that include CPSs, game theory, and network theory. Optimal strategies in the presence of malicious attacks can be designed and analyzed, starting with passive resilience methods, before considering more complex dynamic and adaptive defenses based on reflexive and deliberative methods.

RESILIENT MACHINE LEARNING IN CPSs

CPSs increasingly make use of machine learning (ML) to enable autonomous or semiautonomous decision making in complex environments as well as to improve security and resilience. However, ML techniques themselves can exhibit vulnerabilities to attacks that manipulate observations of the environment or data set used to train the ML models, leading learning algorithms astray with potentially catastrophic consequences.[3]

CPS vulnerabilities can be mitigated by intrusion detection systems (IDSs) that determine whether a particular pattern is caused by normal operation or malicious attacks.[4] IDSs can be constructed using either unsupervised anomaly detection or supervised techniques in which past known attacks, or synthetic cyberattacks, are used in addition to observed normal behavior. The main challenge in CPSs is that the adversary can modify the system behavior to appear similar to normal and, at the same time, compromise a subset of sensors so that the attack cannot be detected. Other novel attacks are also possible. Moreover, ML is increasingly used for learning control behavior based on observations from extensive experiments. For example, in autonomous driving, streaming visual data are used to determine the state of the environment and map observations to steering and speed control actions. A major concern is that the visual data may be corrupted by an adversary, for example, by placing carefully crafted stickers on road markings and signs to induce catastrophic mistakes.[5]

Central problems include the development of foundations for the vulnerability assessment of ML algorithms used in CPSs and algorithms for the identification and analysis of exploitable vulnerabilities. Adversarial models of attacks involving cyber and physical components are also an important problem and must consider coupled attacks. These can involve integrity or denial-of-service attacks on sensor measurements and evasion attacks that minimally modify adversarial behavior to make it appear more benign. Another goal is to develop the algorithmic foundations of resilient ML in CPSs. Game theory methods, for example, are promising for coupled sensor selection, feature selection, and iterative adversarial retraining. Effective attacks are generated using the models developed and then added to training data to introduce examples that the ML approach must correctly categorize as malicious. Resilient ML algorithms in the context of poisoning attacks on training data, especially involving integrity attacks on sensors, can be developed potentially by leveraging redundancy of information provided by different sensors.

METRICS-DRIVEN, SIMULATION-BASED EVALUATION

Heterogeneity and the richness of interactions among components are the key challenges for evaluating security and resilience in CPSs. For some approaches, analytical methods are available for certain attack classes. For example, passivity-based design can be used to analyze the resilience of stability in the presence of denial-of-service attacks.[6] However, this is an exception, and simulation/emulation-based evaluation is typically the main option. In general, what makes the evaluation of security and resilience challenging is its context dependence. Evaluation makes sense only for a well-defined property against well-defined attack classes in a given environment.

Simulation-based experiments are an integral part of developing a systematic body of knowledge with both strong theoretical and empirical underpinnings to inform the engineering of a CPS that can resist cyberattacks. Simulation-based evaluation must use metrics that quantify the degradation of selected properties of system-level behavior in different mission contexts. However, the overall CPS behavior emerges from interactions between multiple heterogeneous abstraction layers. As illustrated in Figure 1, attacks can be deployed in different system layers, and their effects can propagate in the implementation hierarchy. The basic tenet of simulation-based experiments is to build and evaluate executable models for capturing such interactions. Important details about the execution of physical or cyberattacks cannot be ignored. Such details include the formation of network packets, routing information, or operating system-level behavior of individual nodes. This requirement is in conflict with providing a truly generic, technology neutral, and executable adversarial language that captures only abstract high-level domain concepts since the simulation of physical and cyberattacks relies on the details of the implementation infrastructure. The consequences are significant not only for the necessary capabilities of the simulation testbed but also in terms of designing and configuring experiments. The selected system properties and attack types strongly influence the levels of abstraction to be used. In some cases, the required level of simulation fidelity may not be feasible, and using hardware-in-the-loop testbeds is the only possibility.

A CPS modeling and integration platform has been developed for the experimental evaluation of resilient system design in traffic control systems.[7] The evaluation of resilience is based on attacker–defender games using simulations of sufficient fidelity. The platform integrates 1) realistic models of cyber and physical components and their interactions, 2) cyberattack models that focus on the impact of attacks to CPS behavior and operation, and 3) operational scenarios that can be used for evaluation of cybersecurity risks. The heterogeneity required for modeling and simulation of CPSs is addressed by the model, tool, and execution integration platforms. The model integration platform incorporates theories, methods, and tools for formally specifying domain-specific modeling languages, metamodels, and model transformations. The tool integration platform includes an integration framework for simulators based on High Level Architecture 0.8. The execution integration platform includes a range of tools for cloud- or desktop-based deployment. Using the integration platforms, toolkits are used for the rapid configuration and integration of domain-specific CPS design studios, simulators, and analysis tools.[9] An important conclusion of this research has been that semantically rigorous composition of heterogeneous component models is feasible; however, it can quickly result in model sizes that cannot be simulated or formally analyzed due to their excessive complexity.

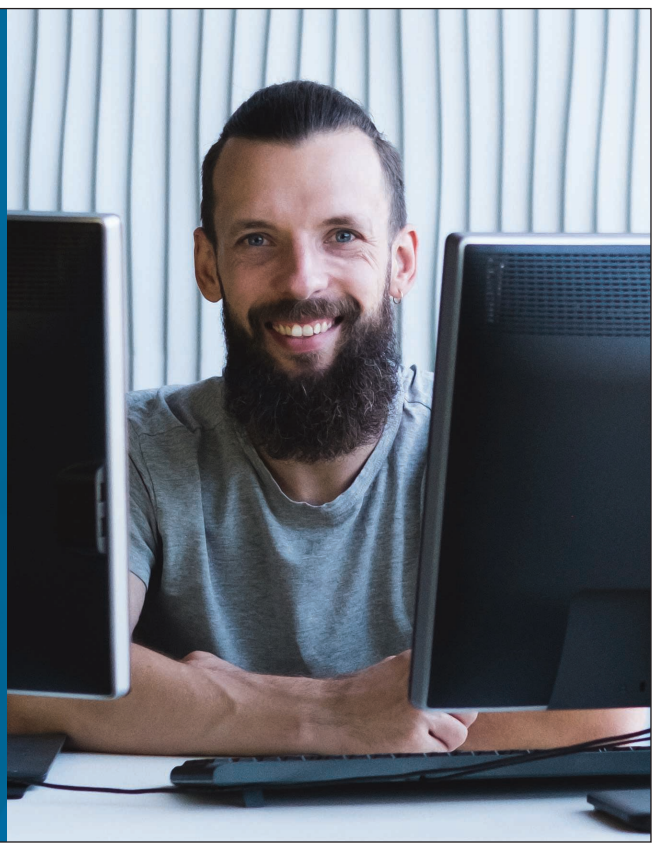
A systems science of secure and resilient CPSs brings together interdisciplinary research with the goal of identifying, exploring, and understanding patterns of complexity that cross disciplines and application domains. The objective is to develop a systematic body of knowledge with strong theoretical and empirical underpinnings and inform the engineering of secure and resilient systems that can resist not only known but also unanticipated attacks.

REFERENCES

1. K. Koscher et al., "Experimental security analysis of a modern automobile," in *Proc. 2010 IEEE Symp. Security and Privacy*, Berkeley/Oakland, CA, pp. 447–462. doi: 10.1109/SP.2010.34.

2. J. Giraldo, E. Sarkar, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in cyber-physical systems: A survey of surveys," *IEEE Des. Test.*, vol. 34, no. 4, pp. 7–17, 2017. doi: 10.1109/MDAT.2017.2709310.
3. N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proc. 2016 IEEE European Symp. Security and Privacy (EuroS&P)*, Saarbrucken, pp. 372–387. doi: 10.1109/EuroSP.2016.36.
4. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2016. doi: 10.1109/COMST.2015.2494502.
5. K. Eykholt et al., "Robust physical-world attacks on deep learning visual classification," in *Proc. 2018 IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, June 18–22, 2018, pp. 1625–1634. doi: 10.1109/CVPR.2018.00175.
6. J. Sztipanovits et al., "Toward a science of cyber–physical system integration," *Proc. IEEE*, vol. 100, no. 1, pp. 29–44, 2012. doi: 10.1109/JPROC.2011.2161529.
7. X. Koutsoukos et al., "SURE: A modeling and simulation integration platform for evaluation of secure and resilient cyber–physical systems," *Proc. IEEE*, vol. 106, no. 1, pp. 93–112, 2018. doi: 10.1109/JPROC.2017.2731741.
8. *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA): Framework and Rules*, IEEE Standard 1516, 2010.
9. J. Sztipanovits, T. Bapty, X. Koutsoukos, Z. Lattmann, S. Neema, and E. Jackson, "Model and tool integration platforms for cyber–physical system design," *Proc. IEEE*, vol. 106, no. 9, pp. 1501–1526, 2018. doi: 10.1109/JPROC.2018.2838530.

XENOFON KOUTSOUKOS is a professor of computer science, computer engineering, and electrical engineering at Vanderbilt University, Nashville, Tennessee. He is a Fellow of the IEEE. Contact him at Xenofon.Koutsoukos@vanderbilt.edu.



Digital Object Identifier 10.1109/MC.2020.2974643