





# Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis

Nelson H. Carreras Guzman<sup>1,2</sup>  | Morten Wied<sup>1,2</sup>  | Igor Kozine<sup>1</sup>  |  
Mary Ann Lundteigen<sup>2</sup> 

<sup>1</sup>Engineering Systems Group, Department of Technology, Management and Economics, Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway

## Correspondence

Nelson H. Carreras Guzman, Engineering Systems Group, Department of Technology, Management and Economics, Diplomvej, Building 371, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark.  
Email: nelca@dtu.dk

## Abstract

Many safety-related systems are evolving into cyber-physical systems (CPSs), integrating information technologies in their control architectures and modifying the interactions among automation and human operators. Particularly, a promising potential exists for enhanced efficiency and safety in applications such as autonomous transportation systems, control systems in critical infrastructures, smart manufacturing and process plants, robotics, and smart medical devices, among others. However, the modern features of CPSs are ambiguous for system designers and risk analysts, especially considering the role of humans and the interactions between safety and security. The sources of safety risks are not restricted to accidental failures and errors anymore. Indeed, cybersecurity attacks can now cascade into safety risks leading to physical harm to the system and its environment. These new challenges demand system engineers and risk analysts to understand the security vulnerabilities existing in CPS features and their dependencies with physical processes. Therefore, this paper (a) examines the key features of CPSs and their relation with other system types; (b) defines the dependencies between levels of automation and human roles in CPSs from a systems engineering perspective; and (c) applies systems thinking to describe a multi-layered diagrammatic representation of CPSs for combined safety and security risk analysis, demonstrating an application in the maritime sector to analyze an autonomous surface vehicle.

## KEYWORDS

automation, cyber-physical systems (CPSs), human factors, Internet of things (IoT), safety and security risks, systems engineering

## 1 | INTRODUCTION

The innovation in cyber-physical systems (CPSs) opens a rising field of multi-disciplinary cooperation, linking computer science and control theory with several engineering areas, natural sciences, and medicine.<sup>1</sup> Increasingly, CPSs are improving performance, productivity, and energy efficiency in the control of physical processes. Researchers and practitioners are designing and prototyping autonomous vehicles (AVs) with higher levels of automation and connectivity.<sup>2</sup> Similarly, the healthcare sector is developing novel medical applications to better support and treat patients, including autonomous implantable devices and system architectures for monitoring patients in hospitals

or at home.<sup>3</sup> Other relevant CPS applications include industrial control systems (ICSs) in manufacturing and process plants, robotics, control systems in critical infrastructures providing essential services to communities<sup>4</sup> (eg, smart grids, water and wastewater systems), and autonomous military defense missiles, among others.

Considering the promising developments and the critical applications of CPSs, government agencies and industrial partnerships regard the research efforts in CPSs as a priority.<sup>5</sup> Consequently, publications in the field of CPSs have experienced a positive exponential rate in annual publications since Hellen Gill coined the term in 2006 at the National Science Foundation (NSF) of the United States.<sup>6,7</sup>

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Systems Engineering* published by Wiley Periodicals, Inc.

However, some researchers acknowledge the challenge to provide an exact conceptualization of CPSs due to the broadness of the term.<sup>8</sup> As a result, the current conceptualizations and representations of CPSs do not properly frame its key features, that is, the essential components and the interactions present in this class of systems.

Furthermore, the relations between levels of automation and human supervision are ambiguous in CPSs. For instance, the NSF defines CPSs as “engineered systems that are built from, and depend upon, the seamless integration of computation and physical components.”<sup>9</sup> In similar terms, Rajkumar et al.<sup>1</sup> characterized CPSs as “physical and engineered systems whose operations are monitored, controlled, coordinated, and integrated by a computing and communication core.” In general, these and other definitions stress the integration of computers to control physical components. According to Alur,<sup>5</sup> this cyber-physical integration arises from sensors and actuators reacting to the physical world. Yet, these definitions tend to assume that CPSs are autonomous systems controlling a set of technical components. In doing so, there is a risk of overlooking or failing to distinguish the vital and evolving roles of humans in the control architectures of CPSs, which are necessary features to assess the safety and security of the system without fuzzy interpretations. Therefore, we stress the need to conceive CPSs as a particular type of socio-technical systems characterized by some new and enhanced key features.<sup>10–12</sup>

Increasingly, CPSs are exposed to security attacks, including intentional cyber threats that can go beyond the information domain and “cascade” into physical hazards in the energy domain. The Stuxnet attack in 2010 to a nuclear facility clearly evidenced this case in reality,<sup>13</sup> while recently perpetrated cyber-attacks mentioned in this paper show the increasing need for cybersecurity in safety-critical CPSs.

The sources of safety risks are not restricted to component failures and accidents anymore. **In CPSs, safety is an emergent property that does not necessarily improve solely by enhancing the reliability of individual components or software.**<sup>14</sup> As a result, risk analysts working on multiple CPS applications require an understanding of the complex interactions and security vulnerabilities existing in general CPS features and their potential to influence safety. Although several methods in the literature have attempted a safety and security analysis integration,<sup>15–18</sup> researchers have paid little attention to providing **a comprehensive system representation of CPSs for designers and risk analysts to visualize the relevant features of the system.**

As Clements accurately affirmed: “we never analyze a system—we analyze only a conceptual model of the system.”<sup>19</sup> For example, practitioners widely rely on piping and instrumentation diagrams (P&IDs) among the system representations to conduct hazard identification in the process industry.<sup>19</sup> Similarly, system and software engineers usually rely on models such as functional block diagrams and Unified Modeling Language (UML) diagrams to represent the software architecture of computer systems and conduct threat analysis.<sup>20</sup> Because these and other representations are not tailored to include the complex interactions in CPSs and their related risks, the field of **safety analysis requires a new systems engineering framework that includes the complex dependencies and the security challenges of CPSs.**<sup>21,22</sup>

This paper addresses the following three research questions.

- (1) Which are the key features of CPSs and their relation with other system types?
- (2) What levels of automation and human control interactions challenge the design of CPSs?
- (3) How can system designers and risk analysts describe the features of CPSs in a comprehensive representation for safety and security analysis?

For each question, we discuss the implications for safety and security risk analysis using recent historical incidents and describing the technologies and system architectures of several CPS applications.

This paper is organized as follows. Section 2 examines the key features that define CPSs as a class of systems, providing an explicit conceptualization of CPSs with key and accessory features and their compatibility with safety and security issues in recent incidents. Section 3 analyzes the levels of automation and the roles of humans in CPSs, highlighting their important repercussions for safety and security risk analysis. Section 4 integrates these previous considerations and applies systems thinking to describe a multi-layered diagrammatic representation of CPSs for safety and security risk analysis. Section 5 demonstrates the suitability of this representation in a case in the maritime sector, specifically as a framework to analyze a real autonomous surface vehicle (ASV). Finally, Section 6 concludes and opens the field for future research in safety and security risk analysis of CPSs.

## 2 | THE EMERGENCE OF CPS: TWO PERSPECTIVES FOR DERIVATION OF KEY FEATURES

A widespread definition of a CPS is the “integration of computation and physical processes.”<sup>23</sup> Nevertheless, the broadness of this and other definitions may obscure the identification of the key features of CPSs, that is, the common characteristics that proof the utility for grouping this wide set of systems into a common class.

When examining CPS applications, one could question the benefits of conceptualizing such a wide set of applications (eg, autonomous vehicles, smart grids, robotics, cutting-edge ICSs, smart medical devices, and military defense systems) in a common class as CPSs. Indeed, this conceptualization would be useful only if it provided practical insights and facilitated the solution of common issues in these applications.

For example, the class of system “car” is useful to provide safety standards and design guidance to different manufacturers, even if they use different technologies and provide accessory features beyond road driving. Furthermore, despite cars being considerably different when compared to other vehicles (eg, motorcycles, trucks, bicycles), one can group all these systems as “road vehicles” to generate common infrastructure and traffic regulations.

Considering the CPS class, Lee<sup>23,24</sup> identified a series of foundational challenges in the abstractions used in computation. He stressed

the need of computer systems to fit the timing requirements of CPSs, that is, concurrent and real-time calculations in networked systems interacting with the physical world.

Focusing on the principles of design, modeling, and verification of the computational components and their integration, Alur<sup>5</sup> proposed a set of **key features of CPSs. Particularly, he mentions reactive computations, concurrency, feedback control, real-time computation, and safety-critical applications.** Whereas this set of features is a useful starting point to categorize CPSs, we argue the need to include the role of humans<sup>25</sup> in CPS design architectures as a key feature of CPSs with safety and security implications. Moreover, we complement Alur's conceptualization with an analysis of other rising systems and paradigms associated—but not identical—to CPSs.

Acknowledging the broadness and fuzziness of the CPS field, Gunes et al.<sup>26</sup> presented a comprehensive survey comparing CPSs to related research fields and concepts such as the Internet of Things (IoT), Machine-to-Machine (M2M) communications, and mechatronics, among others. However, they did not identify explicitly the key features of CPSs for the context of their safety and security challenges, that is, for the protection of CPSs goals against both unintentional and deliberate sources of risk potentially impacting the system or its environment.<sup>27,28</sup>

In this section, we examine the features of CPSs comparing two perspectives of antecedents, applications, and trends for future developments. The first perspective is a well-known approach in the literature, starting the evolution of CPSs from embedded systems (ESs).<sup>5,7,8,29</sup> Still, we introduce how this perspective is also associated with the related field of the IoT, blurring the distinctions between CPSs and the IoT. The second perspective opens a wider landscape of CPSs not necessarily rooted in ESs. Instead, this perspective considers the evolution of control systems in industrial processes and manufacturing leading to CPSs.<sup>6,10,30–32</sup> Within this second perspective, even if there is a tendency to embed the control devices inside the physical components,<sup>31,33</sup> we argue that these features are not essential to define CPSs. By synthesizing the two perspectives previously mentioned, in this section, we define the key and accessory features of CPSs, stressing the need to include explicitly the roles of human operators in CPSs. We finally discuss the association of these features with real cases of safety and security issues in CPSs in recent years.

## 2.1 | First perspective: From ESs to CPSs (and the IoT)

Commonly, the literature considers CPSs as an upgraded stage of ESs.<sup>5,34</sup> In simple terms, ESs are small computers that are not visible to the users. Their origins can be traced back to the 1970s,<sup>6</sup> consisting of “hardware and software integrated within a mechanical or an electrical system designed for a specific purpose”.<sup>5</sup> They are widely implemented in consumer electronics, for example, TVs, digital cameras, smartphones, washing machines, and microwaves. Furthermore, ESs are used in safety-critical applications performing distinct tasks, usually operating in isolated configurations without integration with other real-time control functions.

In contrast to general-purpose computers and industrial controllers, ESs are restricted by their smaller sizes, requiring high levels of design efficiency. Namely, according to Marwedel,<sup>29</sup> ESs should be:

- Energy efficient: considering limited power sources.
- Run-time efficient: avoiding excessive computational time execution and use of memory, energy, and other limited resources.
- Small in code size: considering limited memory size in embedded microcontrollers.
- Lightweight: as they are incorporated into portable physical devices, whose function might be affected by additional weight.
- Low-cost: to achieve cost-effective applications compared to other alternatives in the market.

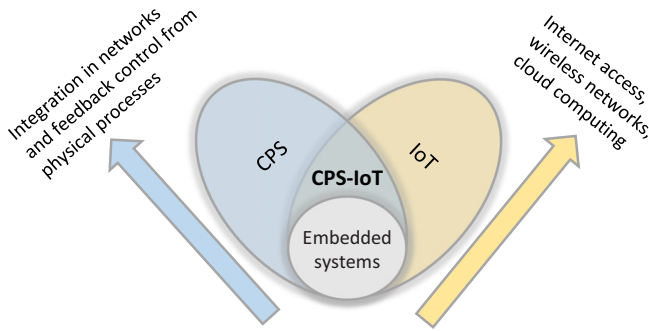
In many cases, ESs operate in open control loops, that is, without incorporating a feedback from the physical processes. This is the case of many consumer electronic goods, being a washing machine a typical example. Moreover, some ESs also operate in open loops in some safety-critical applications, particularly those that do not depend on computers closing the control loops. Indeed, ESs could be used solely on sensor devices, providing data to human operators or to application platforms as a service. Moreover, designers have traditionally conceived ESs in isolation, performing a particular function independent of other ESs and of their environment.

From this perspective, the CPS concept is a paradigm shift for the ESs community. A paradigm shift is a fundamental change in basic concepts describing a scientific discipline. In our case, the aforementioned features contrasting ESs with general-purpose computers are no longer the main issues when designing CPSs. Indeed, “in CPSs, embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa.”<sup>6</sup> Thus, the integration of communication networks and feedback loops from physical processes describe the frontier of the shift from ESs to CPSs.

### 2.1.1 | CPSs and the IoT

In parallel, there is a growing interest on the progressive connection of ESs through computer networks, and specifically to the Internet. Enabling technologies, such as low-power wireless networks, communication protocols, and cloud computing, open the possibility for a new range of applications developed from the interaction of devices connected to the Internet. The design paradigm behind these applications is known as the IoT.<sup>35,36</sup> Services such as smart homes and work places, wireless sensor networks (WSN) in urban and rural infrastructures, industrial automation, and smart healthcare are among some relevant fields of deployment.

The term IoT was proposed by Kevin Ashton in 1999,<sup>37</sup> initially stressing the rising capabilities of radio frequency identifiers (RFID) and wireless technologies. Nonetheless, the term IoT diversified to include a wide set of wireless sensor networks (WSNs). This new IoT paradigm gave rise to different visions and related definitions of the



**FIGURE 1** First perspective: CPS and IoT developments from a perspective centered in ESs

concept according to the historical backgrounds and orientations of different communities.<sup>36</sup>

Some communities argue that the IoT is a key foundation that enables the deployment of CPSs.<sup>38</sup> However, we stress that this field of progress in the IoT is related—but not identical nor essential—to the field of CPSs. For IoT applications, a real-time feedback control of physical processes may not be necessary. Instead, many IoT system architectures develop mobile apps or cloud applications as final services,<sup>36,37</sup> using the integration of smart sensors, wireless networks, internet access, and cloud platforms with advanced data analytics. In contrast, the final services in CPSs are physical systems performing real-time control tasks in the physical world.

Some IoT applications provide smart actuator commands from real-time sensor readings. However, these actions are sometimes limited to the activation of information functions (eg, message display, sound notifications) for surveillance, logistics, and monitoring.<sup>36,37</sup> These simple actions are limited to information awareness, while not completing a physical process by themselves. Instead, CPSs perform control actions in a way that alters the new state of the sensor readings and consequently the states in the control loop by actuator commands with physical effects.

Finally, there is some degree of overlapping between the fields of CPSs and the IoT.<sup>34,39</sup> Mainly, some CPS applications are being connected to the Internet to use data-accessing and processing services.<sup>10,40</sup> Thus, we establish a category of CPS-IoT from the intersection of these two fields, namely, those CPSs built from ESs that include Internet connection in their network configurations.

From this analysis, Figure 1 illustrates CPSs and the IoT as different advances in ESs capabilities. Nevertheless, the overlapping CPS-IoT field (also known as IoT-based CPSs<sup>34</sup>) incorporates both set of capabilities in these systems.

## 2.2 | Second perspective: From cybernetics to CPSs

The field of cybernetics established the foundations for engineered feedback control systems interacting with the physical world, even before the revolution in digital computation and network communications.<sup>6</sup> Norbert Wiener opened the field of cybernetics in 1948, from applications in automatic weapon systems expanding

to a wide field of technical systems and even to human behavior and neuroscience.<sup>41</sup>

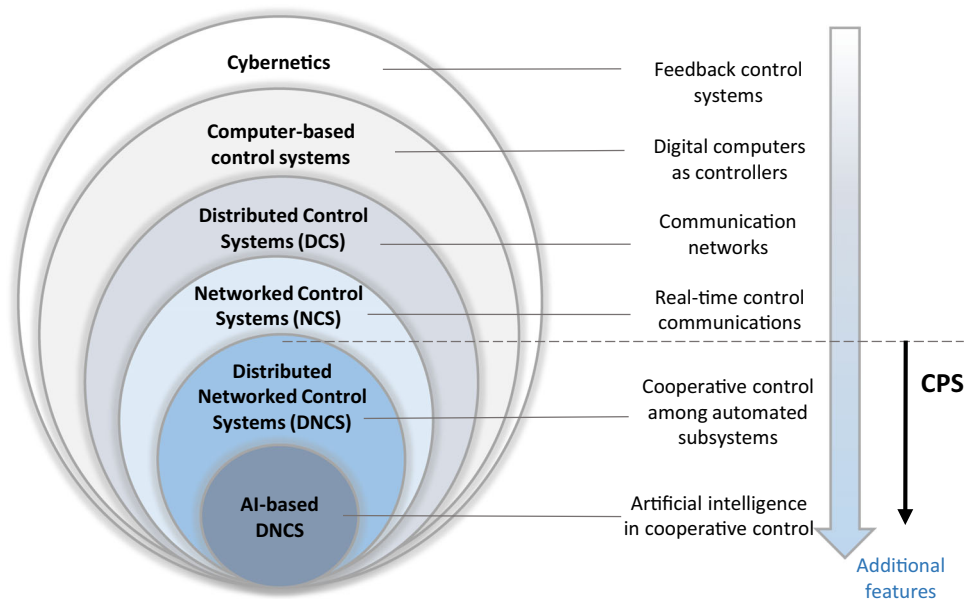
Considering this perspective from the evolution in cybernetics, the notion of CPSs as strictly centered in ESs would be very restrictive. Many control applications tightly coupling cyber and physical processes are composed of programmable controllers not necessarily embedded or hidden into the physical components. In other words, a wide range of industrial devices and operational technology (OT)—for example, supervisory control and data acquisition (SCADA), distributed control systems (DCSs), and programmable logic controllers (PLCs)—are incorporating the operational features of CPSs.<sup>42</sup>

In this sense, a complementary and more comprehensive view of antecedents of CPSs could be traced back to the development of cybernetics in the 1940s as the conjunction of analog computations, communications, and control.<sup>6</sup> These control systems evolved with the introduction of digital computers operating as automatic controllers since the 1960s,<sup>43</sup> when the invention of programmable logic controllers (PLCs) represented a turning point for industrial automation.<sup>33</sup>

Then, improvements in aircraft and industrial process control led to the advent of distributed control systems (DCSs) in the 1970s, enabling remote control operations and a research interest on teleoperation.<sup>30,44</sup> The feedback control was no longer point-to-point. Instead, networked communications were closing the loops, even if each controller node depended only on local information in decentralized configurations.<sup>31</sup> In the late 1980s, networked control systems (NCSs) incorporated real-time communications,<sup>44</sup> introducing technologies such as Ethernet, as well as controller area network (CAN) to connect electronic control units (ECUs) in vehicles.

Nowadays, NCSs are evolving into distributed configurations,<sup>31,44</sup> enabling task coordination and information exchange among automated control subsystems. These distributed networked control systems (DNCSs) are therefore characterized by their capabilities for cooperative control,<sup>31</sup> thus operating as a type of system of systems (SoS).<sup>45</sup> From a system-theoretic perspective, we argue that an SoS integrates formerly independent feedback control loops into an interdependent set of control loops, allowing the realization of cooperative tasks to achieve a higher common goal. Illustrative examples include independent robotic arms holding and rotating together an object in a factory, the steering and braking systems of a car autonomously interacting to avoid a collision, among others. These interactive feedback control loops could also include human supervision and manual control. In the future, these systems could be controlled by artificial intelligence (AI), replacing the preprogrammed algorithms by neural networks and self-improving algorithms.

In summary, Figure 2 illustrates the antecedents and potential future developments of feedback control systems as a progressive evolution of control and communication capabilities. Moreover, in the following subsection, we argue that DNCSs—as conceived in Ge et al.<sup>31</sup>—describe the key features of CPSs, considering the cooperative feedback control capabilities arising from the tight integration of cyber and physical processes. In DNCSs, programmable controllers are no longer isolated in functional—or sometimes even in physical—terms, enabling



**FIGURE 2** Second perspective: Sequential evolution from cybernetics to DNCS as CPS

cooperative tasks between several sensors, controllers, and actuators in dynamic situations.

### 2.3 | The key features of CPS

From the two perspectives described in the previous subsections, we established the antecedents and tendencies for future developments in CPSs. These perspectives are helpful to introduce the key features of CPSs and a comparison with other related system types.

The first perspective centered in ESs is useful to acknowledge the importance of communication networks and feedback loops from the physical processes as two key features distinguishing CPSs from ESs. Even though designers may establish some uncritical CPS functions in open loop configurations, a key feature of CPSs is the capacity to detect the transformations in the physical processes and react in real-time to ensure the functional and safety requirements of the system.

The second perspective, however, expands the notion of CPSs beyond ESs, including OT and industrial devices in wired networks. This second perspective allows the identification of accessory features in CPSs, which are growing tendencies in the field with their specific challenges and prospects. These accessory features can be present in CPSs, but they are not their constituent characteristics. Namely, we consider accessory features:

- Embedded systems (ESs): beyond ESs, our field of CPSs includes systems composed by OT, industrial devices, and general-purpose computers.
- Wireless networks: we include systems networked via wired local area networks (LANs).
- Internet access: in contrast to the IoT paradigm, CPSs can operate without using Internet protocols.

- Fully automated control and AI: CPSs could operate in semi-autonomous configurations and with traditional algorithms.

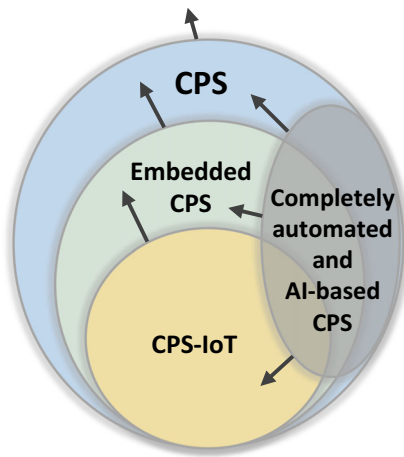
From the previous analysis, we affirm that DNCSs share the relevant features of CPSs. The presence of ESs should not be necessary to characterize CPSs because the specific features of ESs mentioned by Marwedel<sup>29</sup> (eg, size, energy, and memory restrictions) are not the key features in this context. Even considering the increasing tendency to include ESs, wireless networks, and Internet access in DNCSs,<sup>31,44,46</sup> these features do not exclude OT and wired local networks from the domain of CPSs because they provide the same essential function of integrating cyber and physical processes in control systems. Accordingly, wireless networked control systems (WNCs),<sup>44,47</sup> some wireless sensor and actuator networks (WSANs),<sup>31,46</sup> and CPS-IoT applications, are subsets of CPSs equipped with their related accessory features.

Analogously, we consider as accessory (ie, not key) features the fully automatic control capabilities and possible developments in AI-based control in CPSs. Indeed, computers have a major role closing the feedback loops in CPSs, but the human supervision and intermittent intervention are still present and should be considered.<sup>22</sup> In summary, Figure 3 illustrates the field of CPSs with some general subsets describing accessory features in dynamic growth.

Therefore, we define as key features of CPSs the combination of the following:

- (1) Real-time feedback control of physical processes through sensors and actuators
- (2) Cooperative control among networked subsystems, and
- (3) A threshold of automation level where computers close the feedback control loops in (semi)automated tasks, possibly allowing human control in certain cases.





**FIGURE 3** CPS and several subsets with accessory features in dynamic expansion

Subsequently, we conceptualize CPSs as engineered systems that integrate information technologies, real-time control subsystems, physical components, and human operators to influence physical processes by means of cooperative and (semi)automated control functions.

The arguments including DNCs as CPSs are consistent with the literature on safety and security of CPSs. From industrial control in process and manufacturing to embedded and IoT-based, these entire domains share the key features of CPSs and therefore share the potential of cyber threats to disrupt the control system and induce physical harm with safety implications (eg, human injuries, asset damages, and environmental impacts). Recent incidents confirm the rising importance of cybersecurity to ensure safety in diverse CPS applications.

## 2.4 | Compatibility of CPS features in security for safety cases

The realization of cyber threats disrupting SCADA systems and provoking physical consequences could be traced back to the Maroochy water breach in 2000.<sup>48–50</sup> This cyber-attack against the Maroochy Water Services in Australia led to release of one million liters of untreated water into local rivers and parks. A malicious insider (an ex-employee of the system supplier company) used unsecure radio communications to access the control system remotely. Subsequently, the attacker used his knowledge of the system to reconfigure the pumping stations and disrupt the alarms, causing the system to fail in unexpected ways and impeding a rapid response to recover. Even if the water treatment plant was not considered a CPS at the time and may not share all the CPS features, this cyber-attack raised awareness of the security vulnerabilities in critical infrastructures and the potential for physical harm.<sup>51</sup>

Persistently, however, researchers agree on the Stuxnet worm attack to an Iranian nuclear facility in 2010 as the turning point on the physical safety risks exploited by cyber threats in the context of CPSs.<sup>48,49,52–55</sup> The Stuxnet worm entered the system through a USB drive that an operator plugged to a Windows computer. Then, the

worm propagated throughout the SCADA system infecting the PLCs connected to the network. Finally, these PLCs controlling the nuclear centrifuges issued malicious commands to manipulate the rotor speed in ways difficult to detect by the system and the operators, disrupting the physical processes in the nuclear facility and damaging the nuclear reactors.

Neither this nuclear facility nor the Maroochy Water Services plant was composed exclusively of ESs and their system architectures were not completely autonomous. Nonetheless, the control system was not effectively isolated from cyber threats coming from the environment.

Additionally, human operators were unable to respond promptly to the disruptions due to lack of awareness induced by the way the sophisticated cyber threats disrupted or circumvented the alarms. Thus, preventing degradation in situation awareness between the automated system and human operators proves determinant in risk mitigation of CPSs. In this regard, we describe the roles of humans and their implications in CPSs in Section 3.

The vulnerabilities to cyber threats and their potential to cascade into physical harm to human, assets, or the natural environment are shared among a wide set of CPSs (either industrial or embedded). This is the common safety and security issue in autonomous transportation systems, robotics, critical infrastructures, industrial control systems (ICSs) in manufacturing and process plants, and smart medical devices, among others. A series of more recent examples confirms this fact.

Considering other relevant attacks to industrial CPS applications, the German Steel Mill cyber-attack in 2014 caused multiple components of the system to fail, leading to massive physical damages.<sup>56</sup> Using spear phishing e-mails, the attacker gained access to the corporate network and then penetrated into the plant network controlling the physical processes. More recently in 2017, the TRITON malware attack disrupted a petrochemical plant in Saudi Arabia.<sup>57,58</sup> Beyond security concerns of data availability and even operational concerns of continuity in plant operations, this cyber-attack intended to trigger a dangerous explosion in the plant (ie, physical harm). By conducting standard IT intrusion mechanisms, the attackers penetrated into the network and targeted the connected safety instrumented system (SIS).<sup>59</sup> Even though the SIS operated with a proprietary network protocol, enough knowledge of the proprietary system and its connections to general IT networks enable this type of cyber-attacks to target the SIS and induce physical harm.<sup>60</sup>

Embedded CPS applications, such as autonomous vehicles, are also vulnerable to physical harm when subjected to cyber-attacks. Researchers have identified a wide range of cybersecurity vulnerabilities in cars and the potential manipulation of the engine, the steering, and braking system.<sup>61</sup> In 2015, researchers demonstrated how a Jeep model was hacked through Wi-Fi connection, that is, a wireless network providing Internet access.<sup>62</sup> Not only they disrupted the infotainment system, but also they were able to access the CAN bus (the vehicle's wired network) to cut the brakes and shut down the engine while driving on the road. This cyber-physical attack was possible even if this vehicle was not autonomous. With the development of open communications and increasing levels of automation for vehicles, these features become even more important for safety. Beyond cars, similar examples

also show the case of cybersecurity attacks hijacking the control of ships<sup>63,64</sup> and unmanned aerial vehicles (UAVs)<sup>65,66</sup> during operations.

For more historical attacks to CPSs and empirical demonstrations in research environments, Humayed et al<sup>49</sup> described an ample list of attacks to ICSs, smart grids, smart medical devices, and modern cars. They used the description of cross-domain attacks proposed by Yampolskiy et al<sup>48</sup> to discretize the influenced elements (targeted by the attack) from the affected elements (causing the actual damages).

These types of cyber-attacks disrupting physical systems require broadening the scope from security and privacy in CPSs<sup>67</sup> to consider the potential for physical harm and the implications for safety.<sup>27,68</sup> This broader view stresses the need for a combined safety and security risk analysis in CPS, where security and safety goals coexist and require an integration process.<sup>15,69</sup>

### 3 | IMPLICATIONS OF THE ROLE OF HUMANS IN CPSS

In this section, we analyze the implications of the (semi)automated control feature in CPSs, considering the role of humans and their potential influence as sources of safety and security risks.

#### 3.1 | Levels of automation and CPSs

To classify a system as a CPS, the required level of automation is currently unclear.<sup>32</sup> Therefore, we propose as a conceptual threshold, the level where the intended system design assigns the computer the role to close the feedback control loops. In other words, computers have the capacity to gather inputs from sensors and send commands to actuators without the human as an intermediate. Attributing this threshold of automation is not a trivial task because it opens the discussion of the role of humans in CPSs. Particularly, this explicit relationship between a threshold of automation and CPSs as a class of systems serves two relevant purposes.

As first purpose, we delimit the concept of CPSs to the widely agreed domain in the CPSs community, referring to these systems as controlled by a computational core.<sup>1,5–8</sup> The real-time feedback control of physical processes requires hybrid system modeling to integrate the discrete logic of cyber processes with the continuous dynamics of physical processes. Therefore, while some research communities use the concept of CPSs referring to applications in a broader domain, we emphasize this delimitation to frame the key features of CPSs and avoid fuzziness in the concept.

As second purpose, however, we analyze CPSs beyond the automated subsystem using a systems engineering perspective. As Leveson accurately declares: “automation usually does not eliminate humans, but instead raises their tasks to new levels of complexity.”<sup>25</sup> Subsequently, we emphasize the need to analyze the complex interactions between humans in the loop and higher levels of automation.<sup>30</sup> This emphasis is pertinent in CPSs to avoid reducing the system to the technical components and automated functions, but also consider the human roles and their implications in the CPS.<sup>1</sup>

**TABLE 1** Levels of automation, adapted from<sup>30</sup> and<sup>70</sup>

Low	1. The computer offers no assistance: human must take all decisions and actions.
	2. The computer offers a complete set of decision/action alternatives, AND
	3. Narrows the selection down to a few, OR
	4. Suggests one alternative, AND
	5. Executes that suggestion if the human approves, OR
	6. Allows the human a restricted time to veto before automatic execution, OR
	7. Executes automatically, then necessarily informs the human, OR
	8. Informs the human only if asked, OR
	9. Informs the human only if the computer decides.
High	10. The computer decides everything, acts autonomously, ignoring the human.

As shown in Table 1, Parasuraman et al<sup>70</sup> define automation as a continuum of 10 levels where the system increasingly performs functions previously carried out by human operators. These levels of automation usually include conditional connections. If the level ends with an AND, the next level assumes it as an input. For example, in level 5, the system executes a suggestion if the human approves, which assumes the suggested alternative in level 4. Conversely, if the level ends with an OR, the next level imposes a new restrictive constraint. For instance, level 7 executes automatically and then informs the human, while previously level 6 allowed the human a time to veto before automatic execution.

Accordingly, this paper considers a system to be a CPS if it has the capability to operate in level of automation six (6) or higher, in the scale from 1 to 10 shown in Table 1. Namely, as a lower bound or threshold, the computer “allows the human a restricted time to veto before automatic execution.”

Given that this CPS threshold of automation is general in scope but detailed in description, it is suited for extrapolation to other application-based criteria used in different CPSs. For example, the Society of Automotive Engineers (SAE) in their J3016 standard<sup>71</sup> defines six levels of autonomous driving. At level 0, the car has no autonomous capabilities and the human operator is responsible for all aspects of the driving task. Conversely, at level 5, all the driving tasks are managed by the autonomous driving system.

Comparing this criterion from the automotive sector with our CPS threshold of automation, an autonomous vehicle (AV) is a CPS starting from level 2 of the SAE standard. In other words, level 2 of the SAE standard is equal or higher to level 6 in Table 1. At this point, the system executes autonomously and cooperatively the main functions of the car (ie, steering, acceleration, deceleration) reacting to sensor inputs from the physical processes. Nevertheless, the human driver must remain engaged monitoring the environment and should be prepared to intervene physically (eg, turn the steering wheel and push the brakes) if necessary in particular situations.

**TABLE 2** Levels of automation for CPS transportation systems

CPS transport sector	Equivalent level to CPS threshold (level 6 in Table 1)	Level of full automation in sector-based criteria	Reference
Automotive	Level 2 or higher	Level 5	71
Railway	Grade of Automation 2 or higher	Grade of Automation 4	72
Ships	Autonomy level 3 or higher	Autonomy Level 6	73
Aircraft	Level 3 or higher	Level 4	74
UAV	Autonomous control level 4 or higher	Autonomous control level 10	75

Other societies have developed their own criteria for levels of automation in their particular sectors, such as railway,<sup>72</sup> ships,<sup>73</sup> aircraft,<sup>74</sup> unmanned aerial vehicles (UAVs),<sup>75</sup> among others. Table 2 compares these sector-based criteria and assign a threshold in the level of automation to categorize these (semi)autonomous transportation systems as CPSs. For other CPS applications (eg. industrial control systems, smart medical devices), the systems under analysis share similar characteristics when regarded as CPSs.

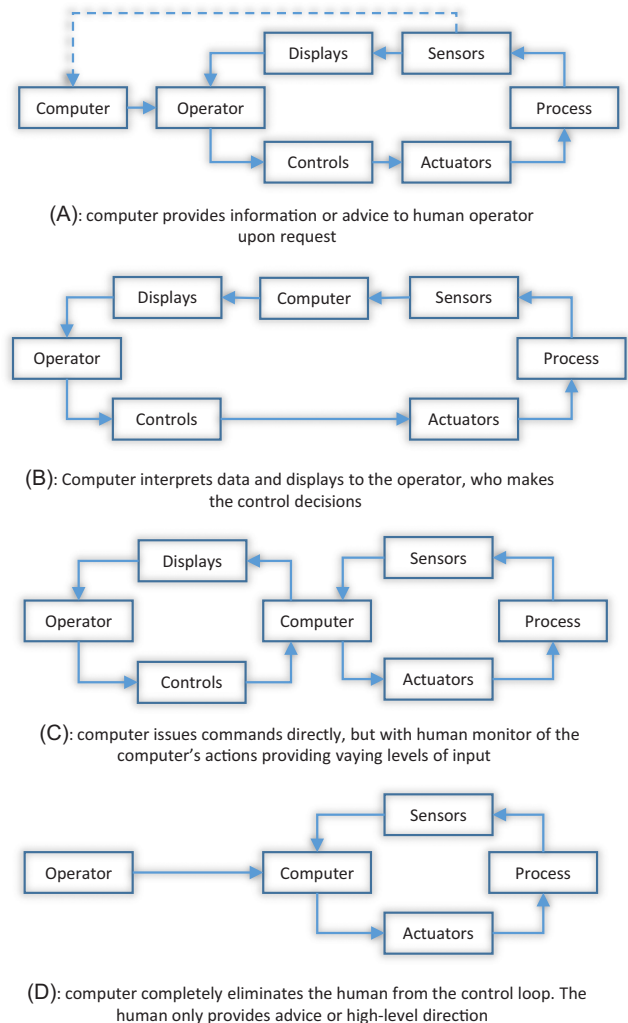
Overall, the CPS threshold of automation (level 6 in Table 1) is equivalent to systems "c" and "d" in the uses of computers in control loops proposed by Leveson,<sup>25</sup> as illustrated in Figure 4. In this figure, the operators are human controllers that (1) collect information of the process under control, (2) use information to make decisions according to models and procedures, and (3) implement control actions to influence the process under control. In systems "a" and "b," this human operator is the only controller in the system. The computer is not a controller because it is not able to provide control actions to the actuators to close the feedback loop with the process under control. In these cases, the computer is only interpreting and displaying information to the operator. The computer becomes an automatic controller in systems "c" and "d," closing partially or even completely the feedback loops. However, the human operator still fulfills the role of supervisor and provides intermittent control actions in some specific cases.

Apparently, the CPS threshold of automation would signify that the system does not close the feedback control loops in real-time, since the human operator has time to veto. Instead, this choice represents the case where the control loops occur in real-time, but the human supervisor can intervene actively if necessary as a feedforward control function (ie, anticipating and reacting beyond feedback corrections).<sup>25</sup>

Additionally, designers can decide that a few control cases are inconvenient as a real-time automatic configuration, considering the complexities involved in the process. Thus, these special control decisions require human decision-making, while the real-time control system directly executes the main functions in normal conditions.

Although one level of automation represents a simplification of a complex control system, this simplification introduces a threshold to frame CPSs with respect to their degree of automated control. In other words, CPSs have the capacity in their control architectures to operate in normal conditions at least in this threshold of automation. Nevertheless, this state of automation can be dynamic, where operators can deliberately take manual control in special cases.

In the future, the evolution in CPSs could lead to fully automatic and adaptable control systems, ceasing to require human supervisory control. Arguably, the advances in AI could entirely substitute the role of

**FIGURE 4** Uses of computers in control loops, adapted from Leveson<sup>25</sup>: This paper considers computer uses (c) and (d) related to the automation threshold for CPS

humans in supervisory control decisions and manual labor.<sup>10,11</sup> However, the current stage of technological development of CPSs is at its infancy<sup>76</sup> and many practitioners recognize the essential need for humans in the loop for the future of autonomous systems.<sup>77</sup> Furthermore, the integration of traditional control algorithms and networked communications with physical processes already pose significant challenges for safety and security, considering their implementation in safety-critical systems.<sup>5,29,78,79</sup>



### 3.2 | Sharing and trading control: Human roles in (semi)automated processes

From a systems engineering perspective, one should not neglect the interactions between humans and technology in automated systems. For example, a designer could be tempted to consider the human as another technical component or a deterministic input-output agent. However, Rasmussen<sup>80</sup> emphasized that human behavior is teleological by nature, that is, operators act according to goal-oriented beliefs predicated on available information. In other words, humans (re)act differently in different situations, in contrast to fixed computer algorithms. In (semi)automated systems, these situations range from repetitive routines with low alertness levels, to unfamiliar tasks under stressful circumstances. We extrapolate this analysis to CPSs, since CPSs are changing the way humans interact with control systems.<sup>1</sup>

As specified by Sheridan,<sup>30</sup> many systems allow for different levels of sharing and trading control between computers and humans. In sharing control, humans and computers perform different control actions in parallel. In other cases, a trading control capability allows for turning complete control to the computer; in case of fully autonomous control, the human only monitors in normal conditions, but can take over partial or total control if necessary.

Therefore, the notion of cooperative control is also possible between computers and humans, expanding the notion of cooperative automatic subsystems presented in DNCs in Section 2.2. In other words, although computers close the feedback control loops in CPSs, humans are in the loop at different levels depending on the system architecture and on the specific circumstance.

This flexibility in automation allows the system to adapt and reduce the level of automation when it is required. On the one hand, this adaptability enhances the system safety under unforeseen circumstances, allowing human operators to deviate the system from prescribed procedures when needed to guarantee safety conditions. On the other hand, this capacity to manipulate the system opens the possibility for erroneous executions, canceling some strengths of system automation such as reliability and predictive performance.

The notion of cooperative control between computers and humans in CPSs is consistent with the theory of distributed situation awareness (DSA).<sup>81</sup> According to DSA, human and nonhuman agents hold situation awareness with different views of the system conditions and with overlapping or complementary goals. In some cases, an agent may compensate the degradation of situation awareness of another agent. This property entails that the system as a whole is the entity that holds all relevant knowledge, whereas different individuals have partial views. However, the partial views of individual agents must be sufficient to perform the tasks assigned to each of them. Even when communication between individual agents is imperfect, they must be able to have awareness of the views of other agents and interpret the information passing through the system.

Several factors compromise the intended human-machine interactions. For example, lack of training, complicated human-machine interfaces (HMIs), lower levels of human alertness, and design constraints, increase the likelihood of accidents due to human-task

mismatches.<sup>25,82</sup> Specially in semi-automated systems with safety-critical scenarios, designers must address the potential reduction of situation awareness in human operators.<sup>83</sup> Conflicting commands with ambiguous privilege protocols between automatic controllers and human controllers could also result in system errors, ranging from degraded performance to economic and safety consequences.

Overall, Leveson<sup>25</sup> concludes that the system benefits from the human presence if designers include the human accountability and responsibilities throughout the design process in a comprehensive way. Accordingly, Parasuraman et al<sup>70</sup> proposed a design framework to evaluate how the types and levels of automation have repercussions for human operators. These considerations apply and should be included in CPSs design for safety.

### 3.3 | Humans as sources of safety and security risk in CPSs

Human roles should be identified and included as potential sources of risk in CPSs. Even if computers close the feedback control loops, humans could still perform complementary roles in cyber processes, such as data insertion, intermittent modifications, and parameter readings, among others.<sup>25</sup> Therefore, we consider humans as crucial actors in CPSs, despite the higher levels of automation incorporated in these systems.

On the one hand, human operators are sources of risk of unintentional motive, that is, with the potential to cause accidents traditionally assessed by safety analysts. In this sense, Taylor<sup>84</sup> described a methodology to assess human error in process plants, covering human error modes as well as latent hazards caused by the system design configuration. On the other hand, both malicious insiders and external cyber-attackers could deliberately disrupt CPSs using acquired knowledge of the system's security vulnerabilities and the dependencies between its system layers.<sup>53,85</sup>

These two sources of risk are different in motive (ie, unintentional and deliberate) and require a comprehensive approach to prevent or mitigate their potential safety-related consequences. Even if these different motives would independently lead to the same harmful consequence, the causal events in each case could require different protection measures in the system. From these considerations, Table 3 summarizes the roles of humans within the system and in the surrounding environments as sources of risks from both unintentional and deliberate motives, that is, from a combined safety and security perspective.

### 3.4 | Humans as prone to safety risks in CPSs

Human safety is also a matter of concern in the physical processes governed by CPSs. For example, new potential **human harm scenarios arise from the capacity of collaborative robots to work alongside humans** in factories, removing the zonal barriers dividing workers and machines and allowing human-robot interaction in physical activities.<sup>86</sup>

Furthermore, accidents involving industrial robots still pose a risk to humans, even when there is apparent separation in their controlled tasks. In December 2018, an accident involving an Amazon's

**TABLE 3** Human roles as sources of risk at different system locations

Risk motives	System		Environment	
	Cyber	Physical	Cyber	Physical
Unintentional	Supervisors using HMI	Physical operators	External operators	Surrounding people
Deliberate	Malicious insiders	Malicious insiders	Hackers	Saboteurs

automated robot punctured a bear repellent spray in a warehouse in New Jersey.<sup>87</sup> After spreading through the warehouse ventilation system, workers became exposed and two dozen of them had to be hospitalized. Moreover, this event is not isolated and has occurred in other facilities.<sup>88</sup>

Other notorious cases for human safety are transportation systems. Recent fatal accidents involved vehicles operating in semi-autonomous mode. The fatal Tesla crash in March 2018 resulted in the death of the driver after the car crashed to a median barrier.<sup>89</sup> During the same month, an Uber in self-driving mode was the first reported crash of an autonomous vehicle killing a pedestrian.<sup>77</sup> These two events resulted from unintentional errors, that is, they did not involve intentional cyber-attacks.

Generalizing these examples to the context of CPSs, safety risks threaten human and assets within the system itself (eg, vehicle drivers, plant workers, patients wearing medical devices). Moreover, safety risks in CPSs also extend beyond the system boundaries and pose concerns to humans, assets, and the natural environment interacting with the system in physical terms. In the next section, we describe a multi-layered diagrammatic representation of CPSs to identify these scenarios and determine their risk sources.

#### 4 | A GENERAL REPRESENTATION OF CPSs FOR COMBINED SAFETY AND SECURITY ANALYSIS

The notion of CPSs is a class of engineered systems grouped by a set of key features. This generalization is a useful framework to analyze these systems according to a common representation, while allowing for the incorporation of their distinctive characteristics within a general framework. As a result, designers, operators, and risk analysts from many disciplines can communicate and collaborate using this common representation as contextual perspective.

Although several methods have attempted a safety and security analysis integration,<sup>15–18</sup> researchers have paid little attention to providing **a comprehensive systems representation of CPSs for designers and risk analysts to visualize the relevant features of the system**. Therefore, the field of safety analysis requires a new systems engineering representation that serves as a basis for a comprehensive safety and security risk analysis method in CPSs.<sup>21,22</sup> This representation should understand the key features of CPSs explained in Section 2 and the evolving roles of humans in automation examined in Section 3. In this context, we consider it necessary to apply systems thinking to encompass the system interactions and feedback loops at different levels.

To facilitate the identification of safety and security risks in a wide range of CPSs, we refine the model of CPS aspects proposed by Humayed et al.<sup>49</sup> In their comprehensive review, they conceived a high-level abstraction of CPSs as the integration of cyber, cyber-physical, and physical aspects. While the physical aspects perform actions in the physical world, the cyber-physical aspects perform reactive computations using sensors and actuators. In other words, the cyber-physical aspects are operational technology (OT) geographically and functionally located in proximity to the physical aspects. Finally, the cyber aspects are higher-level information technology (IT) systems connected to the cyber-physical aspects and, only indirectly, to the physical aspects.

Expanding on this high-level abstraction, we use systems thinking to represent the elements and interconnections of CPSs. According to Arnold and Wade,<sup>90</sup> a valid systems representation grounded on systems thinking should:

- Identify interconnections
- Identify feedback loops and indicate their impact on system behavior (e.g. impact on emergent properties such as safety)
- Illustrate system structure
- Differentiate types of flows and variables
- Identify nonlinear relationships
- Include nonlinear behaviors
- Define the scope to manage complexity by modeling the systems conceptually
- Recognize the system at different scales

Subsequently, we conceive the cyber, cyber-physical, and physical aspects as technologies and entities responsible for the execution of process types in a cooperative multi-layered system. Thus, these processes can be diagrammatically located in layers of the system, controlling particular sets of information and energy flows. In the next paragraphs, we define information and energy flows.

##### 4.1 | Information and energy flows in CPS process types

Information flows are transmissions of information required to achieve the functional goals of the system in the form of computations and communications. These flows include the interactions between operators and technologies through HMIs. In general, the cyber and cyber-physical aspects receive, process, confine, and transmit information flows.

Energy flows are transmissions of energy or matter required to achieve the functional goals of the system in the form of physical work. These flows comprise the physical interactions between operators and physical components through physical interfaces, such as steering wheels and valves, operating machinery using mechanical systems, chemical and biological processes, among others. In general, the physical aspects receive, process, confine, and transmit energy flows. The concept of energy flows excludes the energies used to transmit information flows. Thus, electric signals, electromagnetic waves and other energies involved belong to the domain of information flows as their channels of transmission.

In this sense, we define cyber processes as the uses of IT to control information flows as immediate goal (eg, obtain, store, compute, and transmit). Thus, hardware devices, communication channels, human supervisors, and other physical entities perform cyber processes if their immediate goal is to control information. Conversely, we conceive physical processes as the uses of components different from IT (eg, mechanical, electrical, chemical, and biological) to transform, confine, and control energy flows as immediate goal. At the interface between cyber and physical, cyber-physical processes are particular forms of cyber processes interacting directly and in real-time with the physical ones through feedback loops. Particularly, control mechanisms through sensors, real-time communications, programmable controllers, and actuators compose this category of cyber-physical processes.

From this conception of process types, it becomes explicit that energy flows involved in physical processes are the direct sources of safety hazards (ie, potential physical harm) to humans, assets, and the natural environment. This association between the performance of physical components, the physical human interactions (eg, operation, maintenance), and the energy flows controlled by the system are the traditional focus of safety analysis in physical systems.

However, CPSs integrate cyber-physical processes to control the physical processes through information flows. Consequently, real-time computations and communications are supplanting the human from the control functions, partially removing the human from the physical interface with the system. Moreover, CPS architectures conceive cyber-physical processes as close as possible to the physical ones in both spatial and functional terms, avoiding high latency issues when safety concerns demand a real-time response. For this same reason, real-time control systems (and not human operators) are usually in charge of these critical functions. Traditionally, the potential failures in hardware (HW) and software (SW) leading to accidents in these control systems are the subject of functional safety of programmable electronic systems (PES).<sup>91</sup> Nevertheless, the incorporation of commercial-off-the-shelf (COTS) software and hardware, open-source communications, and standard protocols are introducing vulnerabilities to cyber threats, making security issues a path toward safety risks in the physical processes.

Additionally, CPSs integrate cyber processes on top of the cyber-physical ones. Usually, these information flows are subject to supervision and monitoring in control centers through dedicated HW and SW. Thus, the human role appears as a monitoring and control agent

in those cases where the system architecture did not consider real-time automation. Moreover, the human can access through HMIs the components performing cyber-physical processes to adjust inputs and parameters during different circumstances. Cloud platforms and cloud computing are also possibilities at this level. Furthermore, cyber processes may use different control networks from the cyber-physical ones, thus avoiding data traffic from use cases that do not require real-time processing. Overall, the security of cyber processes are the subject of interest of the cybersecurity field, with emphasis on confidentiality, integrity, and availability. In CPSs, however, cyber-attacks disrupting integrity and availability are the most important security threats leading to safety risks,<sup>78</sup> especially regarding how the influenced information flows affect physical processes.

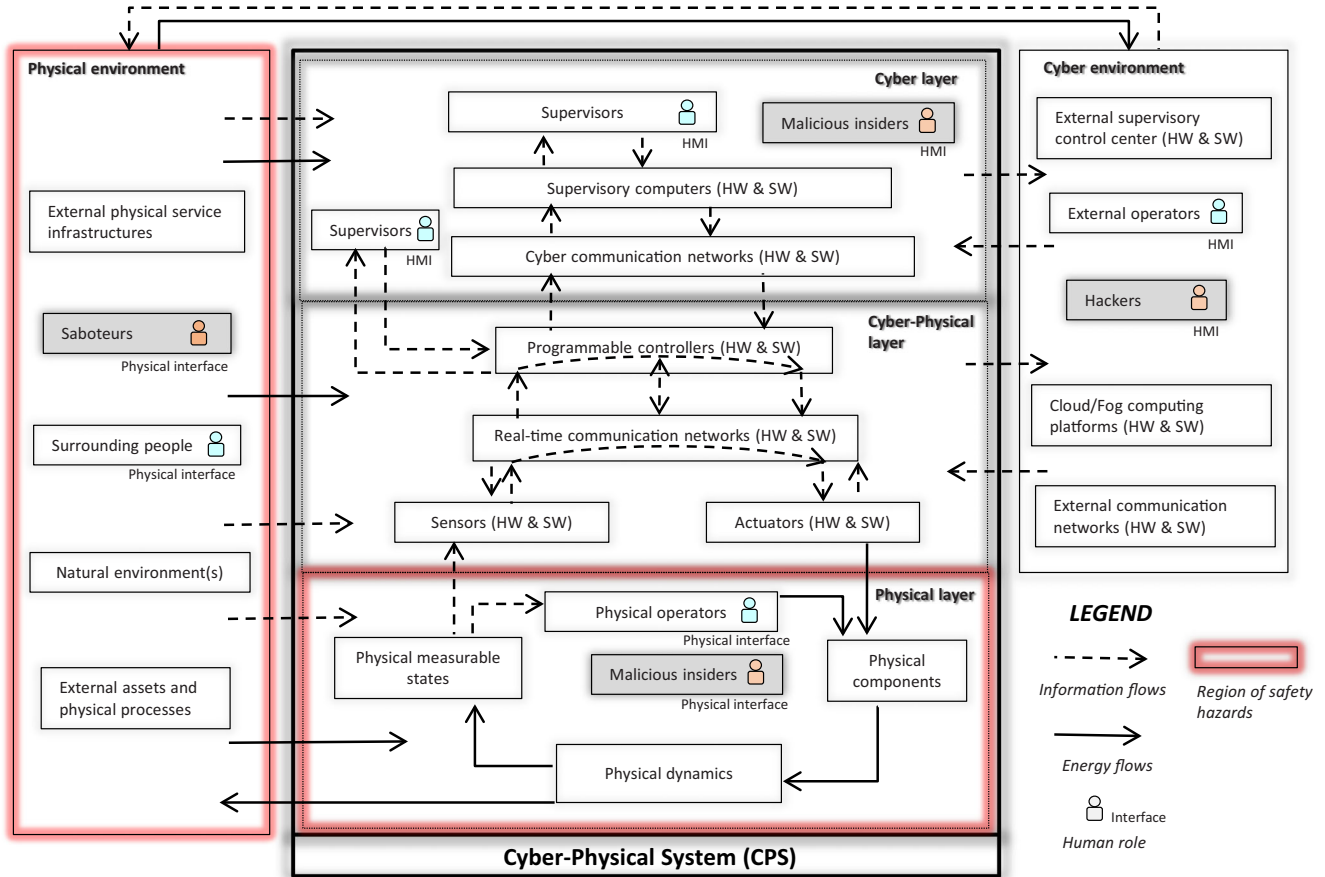
## 4.2 | CPS master diagram: A multi-layered representation for safety and security analysis

To provide a comprehensible representation to professionals from multiple disciplines, we organize the CPS in a hierarchic structure of layers, each layer corresponding to the cyber, cyber-physical, and physical processes. A hierarchic structure is useful to conceptualize the subsystems and their interface interactions,<sup>92,93</sup> in this case from an initial overview at a low level of resolution. Then, we decompose these subsystems in their constituent components with their subsystem interactions, giving a more detailed description of the particular processes.

From the analysis in the previous subsection, in Figure 5, we introduce the CPS master diagram, a diagrammatic multi-layered representation of CPSs as layers of process types. In our hierarchic structure, the lower level of the system is the physical layer, describing the energy flows and the physical interactions to control them. At the middle level, the cyber-physical layer illustrates the real-time information flows to control directly the physical processes through automated feedback control loops. At the top level, the cyber layer presents the information flows for monitoring and supervision. It is possible to visualize the interactions at the system interfaces and derive the mechanisms triggering potential failures across the layers of the system.

The CPS master diagram refines the CPS aspects conceived by Humayed et al in the security context<sup>49</sup> and integrates it with the notion of a control structure as conceptualized by Leveson for a novel accident model in the safety context.<sup>14</sup> The System-Theoretic Accident Model and Processes (STAMP) paradigm conceives safety as a control problem, arising from the interactions between physical processes, automated controllers, and human controllers. Based on STAMP, the System Theoretic Process Analysis (STPA) is a hazard identification technique, which was adapted for a safety and security context as STPA-sec.<sup>94</sup> Despite its significant capabilities to identify accident causes as system interactions beyond individual component failures, in a comprehensive review, Kriaa et al<sup>15</sup> assessed this technique as providing too macroscopic results and not ideal to identify the detailed safety and security interactions.

In this sense, the CPS master diagram provides a more detailed and explicit representation of the processes in CPSs by using the notion of energy and information flows. Furthermore, the CPS master diagram



**FIGURE 5** CPS master diagram: multi-layered representation of CPS and environments with information and energy flows

in Figure 5 includes the physical and cyber environments, that is, those processes that are not under the control of the system stakeholders and that directly influence the state of the system at different layers. Particularly, this representation shows:

- The information and energy flows used by the system to perform different processes in feedback control loops
- The entities and components in charge of providing services to the system, under the control or not of the system stakeholders, and
- The malicious actors that could intentionally disrupt the system at different layers

Note that the physical layer of the system and the physical environment exchange energy flows in both directions, evidencing the region where safety hazards could potentially develop across these interfaces.

This representation serves as a first step toward a combined safety and security risk analysis, providing a generalized diagrammatic illustration of CPS architectures to represent different CPS applications. Note that some blocks and control loops might not be present in specific CPS applications. Nevertheless, we argue that in principle, CPSs possess all three layers in their architectures and usually interact with both the cyber and physical environments.

In the following sections, we describe in detail the features presented in the CPS master diagram.

#### 4.2.1 | Physical layer

In the physical layer, human operators have physical access to the physical components of the system. For example, a human driver manipulates the steering wheel of a vehicle, or an operator manually opens a valve in a process plant. The physical components control a set of physical dynamics, confining energy flows according to the system goals.

In the case of vehicles, they mainly require a control of kinetic energy, while a process plant usually controls a range of energy forms (eg, potential, kinetic, electrical, and chemical). From these physical dynamics, a specific set of measurable quantities provide facts about the state of the system at different timespans. These quantities can be measured by analog sensors or simply perceived by the human operators, who then decide which actions to take to close the feedback control loop in the physical layer.

To conduct these operations, the human should be trained and properly informed of the protocols to follow under different circumstances. Nevertheless, the human capabilities impede in some cases a real-time response, considering the time needed for humans to process information and take an action. In routine tasks, humans might commit errors of distraction, omission, or wrong executions, although with low

probabilities. Conversely, high stress situations and non-routine tasks with reduced time constraints raise the probability of human error.

In terms of security, malicious insiders at the physical layer could use their knowledge of the system to perform dangerous physical manipulations. If the CPS architecture considers this possibility, physical and functional barriers should impede malicious interventions and provide alerts to stop them before leading to hazardous events.

#### 4.2.2 | The interface from the physical to the cyber-physical layer

This interface is the entrance of the physical system into the digital world, where computers and networks control information. Particularly, operational technologies (OT) in the cyber-physical layer incorporate the feedback control functions. As an input, sensors perceive physical quantities from the physical layer and transform them into digital packets. As outputs, actuators are responsible of transforming digital commands into energy flows influencing the physical layer.

#### 4.2.3 | Cyber-physical layer

In the cyber-physical layer, real-time computations and communications take place. This layer is the entrance of the system into the digital world of computers and communication networks, but specifically to those processes requiring real-time response to control directly the physical processes. These processes in the digital world are named cyber-physical processes.

Traditionally in industrial applications and safety-related systems, these cyber-physical processes have been divided in basic process control system (BPCS) and safety instrumented system (SIS), with independent functions and isolated architectures.<sup>95</sup> In contrast, the cyber-physical layer in CPSs increasingly interconnect and integrate the SIS with the BPCS and higher-level computer systems,<sup>96</sup> exposing the system to new safety issues.

In the cyber-physical layer, sensors perceive the measurable quantities from the physical layer and the physical environment, transforming these quantities from analog form into digital form as information flows. These information flows are transmitted through real-time communication networks. They can operate as wired or wireless communications, depending on the system architecture.

In general, these communications should possess some key features. First, the latency must be low enough to guarantee a timely response to the physical layer. Second, they must provide a sufficient quality of service (QoS) to avoid packet losses, operating according to secure protocols. Additionally, these communications are usually made in a dedicated infrastructure, that is, as a separate network from other processes that do not require real-time capabilities, thus avoiding communication jamming and interference.

In some cases, sensors and the actuators are embedded into motes. These motes are embedded systems possessing computation and communication capabilities (eg, a microprocessor and an antenna) integrated with the sensors or actuators. As a result, some cyber-physical processes could close the real-time feedback control loops without recurring to higher-level programmable controllers. Instead, the infor-

mation would flow directly from the sensors through the real-time communication network to reach the actuators.

In most cases, however, sensors and actuators do not have sufficient computation and communication capabilities to close the feedback control loops in cooperative tasks. In these cases, the real-time communication network conveys the information flows from the sensors to higher-level programmable controllers (eg, PLCs, DCS controllers, embedded computers). These higher-level controllers have bigger power sources and more powerful microprocessors to solve complex calculations.

Usually, these controllers are programmed to acquire data from sensors, solve computational algorithms, and finally send commands through the real-time communication network to the actuators. Nevertheless, some complex functions may require multiple programmable controllers to coordinate different actuators in cooperative tasks. Therefore, several controllers could communicate through the real-time network to perform these functions. This is why the information flows between the programmable controllers and the real-time communication network is a two-way arrow in the CPS master diagram in Figure 5.

Although this information needs energy to be transferred and manipulated, the abstraction of information flows stresses the main function of these processes to control (eg, collect, process, and send) information. Subsequently, the energies involved in these processes (eg, electric currents, electromagnetic waves) are means to control information. The same reasoning applies for the cyber layer, as explained in the next paragraphs.

#### 4.2.4 | The interface from the cyber-physical to the cyber layer

In principle, the entire infrastructure in the cyber-physical layer is located at the edge of the physical processes. In this way, the communications do not require long travel distances that could represent a higher latency. Moreover, cyber-physical processes do not incorporate the human controllers. In other words, humans are out of the loop in the cyber-physical layer, considering the real-time capabilities that humans cannot provide through the system using HMIs. These two characteristics (real-time response and human out of the loop) are the main differences between the cyber-physical and the cyber layer.

The transition from the cyber-physical to the cyber layer can materialize in two different ways. The first way is the direct transmission of information flows from the programmable controllers to local human supervisors via HMIs, not requiring real-time processing as monitoring or maintenance functions. The second way is through the digital transmission of information flows from the controllers to the cyber network. In this case, the cyber network transmits this information to describe the state of the system at different time intervals, providing valuable inputs to the supervisory control system and human operators at remote locations. In return, the cyber layer can respond by sending information flows to the programmable controllers or directly to the actuators, allowing trading control capabilities in established cases.

Other particular flows across these layers include the cases where humans edit the parameters of sensors through HMIs. Similarly,



actuators could send information flows about their status to the cyber layer. These two cases go in the opposite direction to the main loops of the system, evidencing the complexity in the dependencies in CPSs between their cyber-physical and cyber layers.

The integration of these cases are the means by which cyber threats (unintentionally or deliberately) disrupting the cyber layer can propagate to the cyber-physical layer.

#### 4.2.5 | Cyber layer

The cyber layer encompasses those processes in the digital world of information technologies (IT) that do not require real-time response, where human operators can perform the role of supervisors. The processes at this layer are cyber processes. Among the related technologies, we include supervisory control and data acquisition (SCADA) systems, HMIs, supervisory computers using cloud platforms for data visualization and parameters adjustment, among others. Subsequently, the human appears at the higher-level of the loop, monitoring and controlling through HMIs.

Therefore, despite being out-of-the-loop at the cyber-physical layer, the CPS master diagram considers the case of human in the loop at the cyber layer. At this level, humans have the capabilities to provide inputs, edit cyber-physical parameters, and even gain full control of the system exploiting the cooperative control capabilities of the system, for example, taking intermittent control of the actuators through remote operation.<sup>25,30</sup>

The connectivity of the cyber layer to the cyber-physical allows malicious insiders to use HMIs to attempt disruptions of the cyber-physical layer. Several protection techniques (eg, isolating privileged execution domains, authentication and access control, and firewalls) and response strategies (eg, intrusion detection) can prevent these cyber security threats by impeding their propagation down through the CPS layers.

Air gaps (ie, network isolation from the cyber environment) are common security measures to prevent the exposure of the safety-related systems to cyber threats. Nevertheless, malicious insiders could also perform cyber-attacks by having physical access to the local cyber network and injecting malware through vulnerable ports. As mentioned in Section 2.4, the Stuxnet worm entered an Iranian nuclear facility via unprotected USB ports and then propagated throughout the network until reaching the PLCs. System designers and operators need to consider these risks in CPSs, reducing the attack surfaces by connecting only the necessary hardware to the networks and providing protocols to disable them if intrusions occur.

#### 4.2.6 | CPS system boundary and the surrounding environments

The CPS master diagram considers a CPS as a system of three interacting layers. However, a complete picture of CPSs should consider that this system is also interacting with its environments. In this context, we draw the boundaries between the system and its environment with respect to the domain of responsibility of the CPS stakeholder.

The system is composed by the cyber, cyber-physical, and physical processes that are within the control of the system stakeholder (eg, the plant or infrastructure managers, vehicle operators, medical device managers). Outside this domain, we subdivide the environments interacting with the system into a cyber and a physical environment. In terms of the CPS master diagram, the cyber environment is only exchanging information flows with the cyber and cyber-physical layers of the system, while the physical environment interacts with all the layers of the system through energy and analog information flows. The following paragraphs explain the characteristics of these environments and their interactions with the CPS.

#### 4.2.7 | Physical environment

**The physical environment is the set of external entities, infrastructures, and natural environments interacting with the CPS in functional terms through energy flows and analog information flows.** Some common examples include external providers of physical resources (eg, electric power, water, gas), external assets and infrastructures (eg, road networks, surrounding vehicles, construction sites, agricultural infrastructure), people physically interacting with the system (eg, pedestrians, passengers, workers, residents in the vicinity of a plant), and the natural environment influencing the system performance and being influenced by the system outputs.

The physical environment influences the CPS via the input of energy flows to the system at all its levels. From a safety and security point of view, both unintentional and deliberate disturbances may arise from the physical environment and disrupt some processes inside the system. Natural hazards (eg, earthquakes, storms, floods, and wildfires), power blackouts, and other physical service interruptions are examples of unintentional disturbances, while malicious manipulation of external infrastructures, bomb explosions, and asset theft are examples of deliberate attacks arising from saboteurs in the physical environment.

In the opposite direction, the CPS can also influence its surrounding physical environment. From a safety point of view, elements of the physical environment may be vulnerable to CPS-driven hazards. Particularly, people, assets, or natural environments located geographically near to the physical layer of the CPS (or describing physical dependencies with CPS functions) may experience losses due to hazardous events arising from within the CPS. In the CPS master diagram, this case arises as an uncontrolled flow of energy going from the physical layer of the system toward the physical environment. In other words, energy outputs of the system (eg, kinetic, chemical, thermal, and radioactive) can become safety hazards to the physical environment when a loss of confinement of these energies occur.<sup>97</sup> These hazards may materialize into physical harm to people (eg, fatalities, severe injuries), asset damage (eg, collisions, fires, explosions), or impacts to the natural environment (eg, pollution, biodiversity loss, and ecosystem degradation).<sup>19</sup> All these safety-related impacts are the subject of study of safety analysis, while other types of losses (eg, financial losses, reputation losses) may also result from these incidents. Loss of control of vehicles or mobile machinery, plant explosions, fires, and toxic

releases are some hazardous events arising from the physical layer with the potential to cause physical harm to the physical environment.

#### 4.2.8 | Cyber environment

In contrast to the physical environment, the cyber environment is the set of external infrastructures and services interacting with the CPS in functional terms through digital information flows. This domain may include a wide range of information and communication services.

A first example is the case of external communication networks, that is, communication systems not controlled by the CPS operators. The cyber and cyber-physical layers of the CPS may operate using cellular communications provided by external vendors, while also these layers can be connected to the internet via internet service providers to exchange relevant information with other systems.

A second example is the case of external control centers, that is, supervisory control functions performed by external service suppliers. Similarly, a combination of internal and external supervisory control systems may be present in the supervisory control architecture.

Finally, a third example is the case external computer servers, namely cloud or fog platforms connected to the CPS. In these cases, some information processing functions may be executed by HW and SW components outside the domain of control of the CPS stakeholders.

Specifically, the architecture of the system could include cyber-physical processes to be executed by fog platforms, providing higher computing power and real-time control functions at affordable costs. Similarly, cloud platforms could provide user-friendly interfaces and computer power to process and display monitoring data. The main difference between a fog platform and a cloud platform resides on the capacity of the fog to provide real-time response to control the physical system. This real-time response is accomplished by allocating the computing resources at the edge of the physical processes (also known as edge computing<sup>35</sup>), thus reducing the processing latency and allowing real-time responses. For this reason, fog platforms could be implemented as external infrastructures assisting the cyber-physical processes of the CPS, while cloud platforms are only recommended to assist cyber processes at higher-levels of supervision and monitoring.

In all these three examples, the CPS stakeholders are not responsible for maintaining and assuring the service continuity of these external systems. Nevertheless, a safety and security analysis in CPSs should include the possibility of both unintended and deliberate deviations in the information flows coming from the cyber environment. These deviations include, for example, the scenarios of service disruption due internet disconnection, SW errors in cloud/fog platforms, or packet losses in communication with the system.

Moreover, the interactions of the CPS with the cyber environment also allows the possibility of targeted cyber-attacks to the specific platforms used by the CPS. These cyber-attacks performed by hackers can disrupt the service availability as Denial of Service (DoS) attacks and compromise the integrity of the information flows as Man in the Middle (MITM) attacks, among others. If attackers have enough knowledge of the CPS architecture and protocols, they could use these types of attacks to penetrate into the cyber-physical layer and provoke damage

all the way through the physical layer of the system and the physical environment.<sup>98</sup>

An appropriate protection of CPSs against cyber-attacks from the cyber environment should include a systematic process of patch management and periodic maintenance of security measures. The relevance of these processes in security were evident after the WannaCry ransomware attack in 2017, which exploited a vulnerability in Windows computers that Microsoft had patched 2 months before. Because unpatched systems were vulnerable to WannaCry, the ransomware encrypted around 300 000 computers in 150 countries and affected critical organizations such as the National Health Service (NHS) in the United Kingdom.<sup>99</sup>

Open communication protocols are increasingly included in CPS architectures, especially in CPS-IoT applications.<sup>100</sup> Despite their short ranges, wireless communication technologies such as Radio-frequency identification (RFID) and Near-field communications (NFC) are not exempt from intentional attacks.<sup>37</sup> Similarly, wireless LAN (WLAN) protocols (such as Bluetooth, Wi-Fi, and others) can contain vulnerabilities, allowing cyber-attackers to tamper the communications from remote stations or even use these networks as attack surfaces to inject malware into the system.

Unaware or manipulated operators could also be the gateways leading to cyber-attacks. Unsafe manipulation of network (eg, plugging infected drives to a workstation, clicking phishing attack links) can lead to the propagation of viruses throughout the cyber layer with potential repercussions to the cyber-physical layer. As a result, both accidental and intentional deviations arising from the cyber environment could disrupt the information flows of the CPS and cascade downstream until becoming safety hazards in the energy flows of the physical layer.

#### 4.2.9 | External interactions between cyber and physical environments

For completeness, the CPS master diagram includes the energy and information flows exchanged between the cyber and physical environments of the system. These interactions illustrate the dependencies that could exist among the different environments. While normally these interactions fall outside the scope of the safety analysis of the system, some critical CPSs could require a deeper analysis of the environmental functions surrounding the system. In this way, the environmental deviations eventually disrupting the system could be traced back to their external causes, following the causal chains across different systems with several stakeholders involved. These cases could be relevant in critical infrastructure protection and other CPSs with regional or national security implications.<sup>101,102</sup>

## 5 | A DEMONSTRATION OF THE CPS MASTER DIAGRAM IN THE MARITIME SECTOR

In this section, we demonstrate an application of the CPS master diagram to represent an ASV. In doing so, we demonstrate the advantages



**FIGURE 6** The Telemetron ASV vessel platform. Courtesy of Maritime Robotics

of conceptualizing the key features of CPSs and using the CPS master diagram for system representation. Finally, we introduce the promising method for comprehensive safety and security risk identification using the CPS master diagram.

In this case, we analyze the Telemetron ASV—Maritime Robotics' Polar Circle 845 Sport vessel<sup>103</sup>—a real scale testbed that incorporates autopilot mode and a collision avoidance system (COLAV). Figure 6 illustrates the vessel driving at sea.

## 5.1 | Conceptualizing an autonomous surface vehicle as a CPS

The Telemetron ASV is equipped with radar sensors and an automatic identification system (AIS), the latter being the integration of a satellite navigation system with an inertial navigation system. In the ASV architecture, a programmable controller on-board reads the inputs from these sensors and processes them according to a control logic, providing the system with the capability to operate in autopilot mode while navigating at sea according to a pre-established route. Moreover, a COLAV is able to detect other vessels in the vicinity, modify the route, and issue the corresponding control commands to avoid collision according to the international regulations for preventing collisions at sea (COLREGS).<sup>103</sup>

In the following paragraphs, we argue that this ASV incorporates the key features of CPSs described in Section 2.

- (1) The real-time feedback control of physical processes through sensors and actuators is fulfilled via the on-board programmable electronic system. Namely, localization and detection sensors give inputs to the programmable controller on-board. In turn, the controller issues commands to the steering and propulsion actuators to drive the vessel in autopilot mode. These actions influence the states of the ASV and the environment, including the maneuvers of surrounding vessels. Finally, sensors detect the new states of the system and the environment, closing the feedback control loops in real-time while driving.
- (2) The cooperative control among networked subsystems consists of the trading control capabilities of the system to change from autonomous to manual mode. Operators can obtain this control in

two ways, according to the system architecture. In the first mode, on-board operators can use a switch to take over physical control of the steering wheel and the electromechanical propulsion system (human control at physical layer). In the second mode, human supervisors using HMIs wired to the controller can use a button in the screen to take control over the vessel and assign steering angles and propulsion speed (human control at cyber layer). Moreover, humans in a remote control workstation could also take control over the vessel as described in the second mode, but in this case transmit the information through wireless communications.

- (3) The threshold of automation level where computers close the feedback control loops in (semi)automated tasks (level 6 or higher in Table 1) is also present. The ASV can navigate at sea in autopilot mode and react to obstacles in the environment through the COLAV system. From a sector-based criteria of autonomy levels (AL),<sup>73</sup> we categorize this system as AL4: human on the loop-operator/supervisory. At this level, "decisions and actions are performed autonomously with human supervision. High impact decisions are implemented in a way to give human operators the opportunity to intercede and over-ride them."

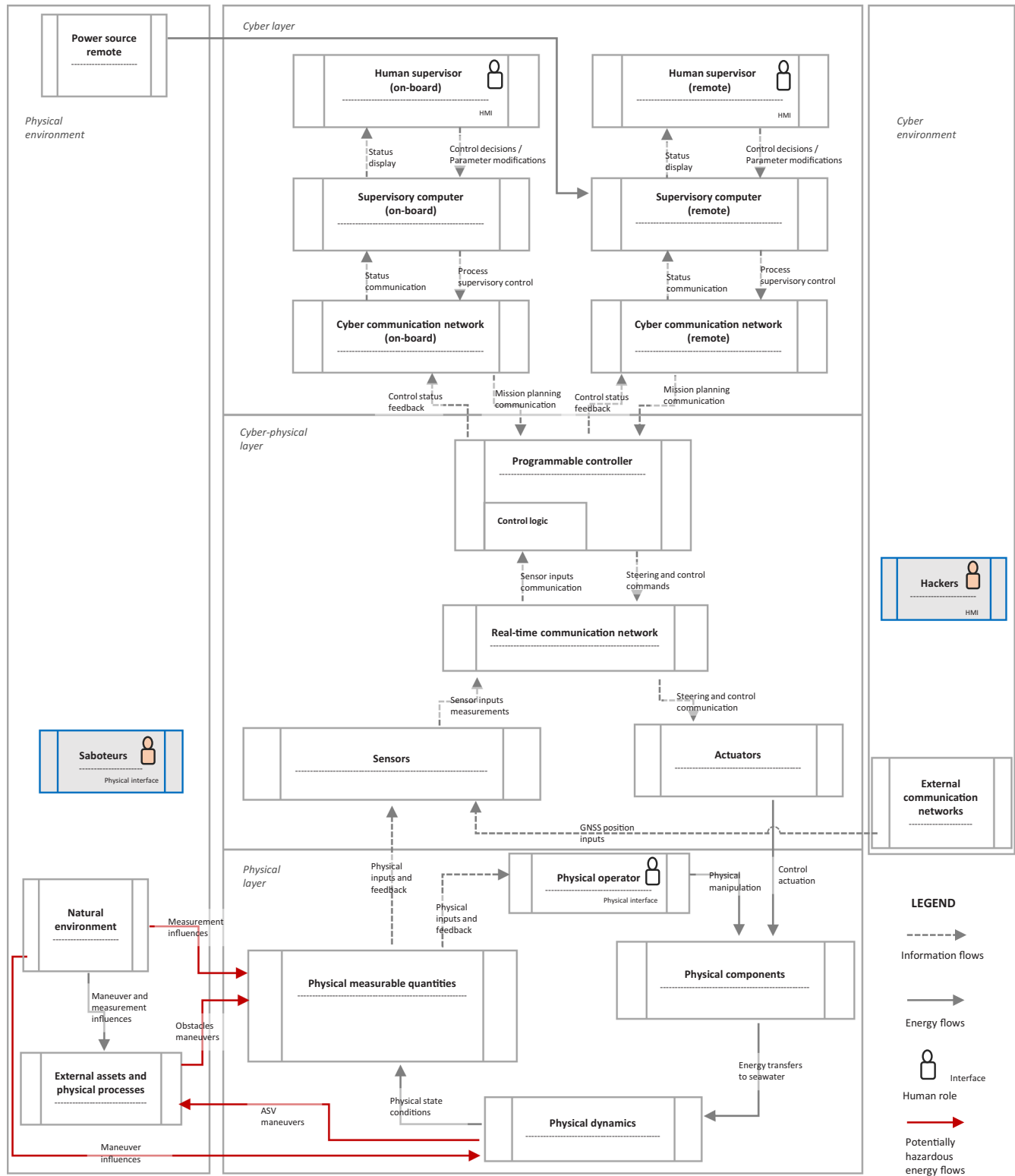
Consequently, in Figure 7, we represent the ASV as a CPS using the CPS master diagram presented in Section 4. This diagrammatic multi-layered representation identifies the information and energy flows and their feedback loop interactions. The CPS master diagram also identifies the potentially hazardous energy flows at the interface between the physical layer and the physical environment, describing uncontrolled flows of kinetic energy that could result in collisions.

Finally, industrial recommended practices such as DNVGL-RP-0496<sup>63</sup> stress the potential of cyber-attacks to penetrate the marine vessels, disrupting the operational technology (OT) of the system, and reaching physical consequences. Therefore, we locate potential attackers at physical environment (saboteur) as well as the cyber environment (hacker). In Appendix A, we provide an expanded version of the CPS master diagram of the ASV, illustrating the specific technologies inside each component block and a selection of types of attacks potentially disrupting the system at different layers.

## 5.2 | A concept for combined safety and security risk analysis: Avoiding physical harm

Using the CPS master diagram defined in this paper as a framework for risk analysis, practitioners from multiple disciplines can apply existing or new risk identification techniques to analyze different CPS applications. As an alternative, in further work we aim at providing a risk identification method for CPSs, conceptualizing the deviation of cyber processes as Uncontrolled Flows of Information (UFoI). These deviations—ranging from unintended incidents to deliberate attacks—are sources of risk to the system at the cyber and cyber-physical layers.

The concept of UFoI refines the Uncontrolled Flow of Energy (UFoE) model proposed in<sup>97</sup> to the field of CPS, considering that cyber, cyber-physical, and physical processes are interdependent and interact with their environments. Therefore, we could model the dependencies



**FIGURE 7** CPS master diagram as system representation of autonomous surface vehicle (ASV)

between safety and security sources of risk leading to physical harm as the cascade of UFoI into UFoE, that is, as Uncontrolled Flows of Information and Energy (UFoI-E).<sup>104</sup> This concept of UFoI-E is com-

patible with the notion of security for safety,<sup>68</sup> where the focus is to enhance safety risk analysis considering the evolving types of physical and cyber-attacks that could lead to physical harm in CPSs.

## 6 | CONCLUSIONS

This paper conceptualized CPSs as engineered systems that integrate information technologies, real-time control subsystems, physical components, and human operators to influence physical processes by means of cooperative and (semi)automated control functions. We identified the key features of CPSs as (1) real-time feedback control of physical processes through sensors and actuators; (2) cooperative control among networked subsystems; and (3) a threshold of automation level where computers close the feedback control loops in (semi)automated tasks, possibly allowing human control in certain cases.

Furthermore, we identified a threshold of automation and its implications for the role of humans in CPSs. This explicit relationship between a threshold of automation and CPSs served two relevant purposes. On the one hand, as widely agreed in the CPSs community, we delimited the scope of the CPSs field to the control systems where computers close feedback control loops automatically and in real-time. On the other hand, we analyzed CPSs beyond the automated subsystem using a systems engineering perspective, examining the complex interactions between humans in the loop and higher levels of automation.

Finally, we integrated the previous discussions and applied systems thinking to introduce the CPS master diagram, a multi-layered diagrammatic representation of CPSs useful to perform risk analysis to a wide range of system applications. The CPS master diagram classified physical, cyber-physical, and physical processes according to the concept of information and energy flows, assisting stakeholders and risk analysts from multiple disciplines in the comprehension of CPSs with their related safety and security considerations to prevent physical harm. We demonstrated the suitability of the CPS master diagram to represent an ASV and to serve as a framework to perform a safety and security risk analysis. In further work, we will integrate the CPS master diagram with the UFol-E concept, generating a method to perform a combined safety and security risk analysis and support responsible innovation in CPS applications.

## ACKNOWLEDGMENTS

We are very grateful for the valuable feedback and useful suggestions of three anonymous reviewers. Their constructive inputs helped us improve this paper.

## ORCID

Nelson H. Carreras Guzman 

<https://orcid.org/0000-0002-8528-1299>

Morten Wied  <https://orcid.org/0000-0002-2034-5490>

Igor Kozine  <https://orcid.org/0000-0002-9698-7211>

Mary Ann Lundteigen  <https://orcid.org/0000-0002-9045-6815>

## REFERENCES

1. Rajkumar R, Lee ILI, Sha LSL, Stankovic J. Cyber-physical systems: the next computing revolution. *Des Autom Conf (DAC)*. 2010 47th ACM/IEEE. 2010:0-5. <https://doi.org/10.1145/1837274.1837461>.
2. Jang C, Sunwoo M, Jo K, Kim J, Kim D. Development of autonomous car—part I: distributed system architecture and development process. *IEEE Trans Ind Electron*. 2014;61(12):7131-7140.
3. Haque SA, Aziz SM, Rahman M. Review of cyber-physical system in healthcare. *Int J Distrib Sens Networks*. 2014;2014. <https://doi.org/10.1155/2014/217415>.
4. United States Department of Homeland Security. NIPP 2013: partnering for critical infrastructure security and resilience. <https://www.dhs.gov/sites/default/files/publications/national-infrastructure-protection-plan-2013-508.pdf>. Accessed May 7, 2019.
5. Alur R. *Principles of Cyber-Physical Systems*. Cambridge, MA: The MIT Press; 2015.
6. Lee EA, Seshia SA. *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*. Cambridge, MA: The MIT Press; 2017.
7. Helen G. A continuing vision: cyber-physical systems. *Fourth Annual Carnegie Mellon Conference on the Electricity Industry Futur*; 2008:1-28.
8. Park KJ, Zheng R, Liu X. Cyber-physical systems: milestones and research challenges. *Comput Commun*. 2012;36(1):1-7.
9. Cyber-Physical Systems (CPS). National Science Foundation. [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=503286](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286). Accessed December 10, 2018.
10. Monostori L, Kádár B, Bauernhansl T, et al. Cyber-physical systems in manufacturing. *CIRP Ann Manuf Technol*. 2016;65(2):621-641.
11. Broy M, Cengarle MV, Geisberger E. Cyber-physical systems: imminent challenges. *Lect Notes Comput Sci*. 2012;7539:1-28.
12. Chatzimichailidou MM, Protopapas A, Dokas IM. Seven issues on distributed situation awareness measurement in complex socio-technical systems. *Complex Syst Des Manag*. 2014:105-117.
13. Langner R. Stuxnet: dissecting a cyberwarfare weapon. *IEEE Secur Priv*. 2011;9(3):49-51.
14. Leveson NG. *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: The MIT Press; 2011.
15. Kriaa S, Pietre-Cambaces L, Bouissou M, Halgand Y. A survey of approaches combining safety and security for industrial control systems. *Reliab Eng Syst Saf*. 2015;139:156-178.
16. Rasputnig C, Opdahl A. Comparing risk identification techniques for safety and security requirements. *J Syst Softw*. 2013;86(4):1124-1151.
17. Bolbot V, Theotokatos G, Bujorianu LM, Boulougouris E. Vulnerabilities and safety assurance methods in cyber-physical systems: a comprehensive review. *Reliab Eng Syst Saf*. 2018;182:179-193.
18. Chockalingam S, Hadziosmanovic D, Pieters W, Teixeira A, van Gelder P. Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications. In: *Critical Information Infrastructures Security. 8th International Workshop, CRITIS 2013. Revised Selected Papers: LNCS 8328*. Vol 8328; 2013. <https://doi.org/10.1007/978-3-319-03964-0>
19. Rausand M. *Risk Assessment: Theory, Methods, and Applications*. Hoboken, NJ: John Wiley & Sons; 2011.
20. Lund MS, Solhaug B, Stølen K. *Model-Driven Risk Analysis: The CORAS Approach*. Berlin, Heidelberg: Springer; 2011.
21. Zio E. The future of risk assessment. *Reliab Eng Syst Saf*. 2018;177:176-190.
22. Sinclair M, Siemieniuch C, Palmer P. The identification of knowledge gaps in the technologies of cyber-physical systems with recommendations for closing these gaps. *Syst Eng*. 2018;22:3-19.
23. Lee EA. Cyber physical systems: design challenges. In: *11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing*. Washington, DC: IEEE Computer Society; 2008:363-369.



24. Lee EA. Cyber-physical systems – are computing foundations adequate? In: *NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap*. Austin, TX; 2006:1–9.
25. Leveson N. *Safeware: System Safety and Computers*. Boston, MA: Addison-Wesley; 1995.
26. Gunes V, Peter S, Givargis T, Vahid F. A survey on concepts, applications, and challenges in cyber-physical systems. *KSII Trans Internet Inf Syst*. 2014;8(12):4242–4268.
27. Pietre-Cambaces L, Bouissou M. Cross-fertilization between safety and security engineering. *Reliab Eng Syst Saf*. 2013;110:110–126.
28. Pietre-Cambaces L, Chaudet C. The SEMA referential framework: avoiding ambiguities in the terms “security” and “safety.” *Int J Crit Infrastruct Prot*. 2010;3(2):55–66.
29. Marwedel P. *Embedded System Design - Embedded Systems Foundations of Cyber-Physical Systems*. Berlin, Germany: Springer; 2011.
30. Sheridan TB. *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: The MIT Press; 1992.
31. Ge X, Yang F, Han QL. Distributed networked control systems: a brief overview. *Inf Sci*. 2017;380:117–131.
32. Wang L, Törnngren M, Onori M. Current status and advancement of cyber-physical systems in manufacturing. *J Manuf Syst*. 2015;37:517–527.
33. Monostori L. Cyber-physical production systems: roots, expectations and R&D challenges. *Proc CIRP*. 2014;17:9–13.
34. Pazzi L, Pellicciari M. From the internet of things to cyber-physical systems: the holonic perspective. *Proc Manuf*. 2017;11:989–995.
35. Al-Fuqaha A, Guizani M. Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun Surv Tutor*. 2015;17(4):2347–2376.
36. Atzori L, Iera A, Morabito G. The Internet of things: a survey. *Comput Networks*. 2010;54:2787–2805.
37. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): a vision, architectural elements, and future directions. *Futur Gener Comput Syst*. 2013;29(7):1645–1660.
38. Carruthers K. Internet of Things and Beyond: Cyber-Physical Systems. *IEEE Internet of Things Newsletter*. <https://iot.ieee.org/newsletter/may-2016/internet-of-things-and-beyond-cyber-physical-systems.html>. Accessed December 12, 2018.
39. Ochoa SF, Fortino G, Di Fatta G. Cyber-physical systems, internet of things and big data. *Futur Gener Comput Syst*. 2017;75:82–84.
40. Stankovic JA. Research directions for the internet of things. *IEEE Internet Things J*. 2014;1(c):3–9.
41. Wiener R. *Cybernetics: Or Control and Communication in the Animal and the Machine*. 2nd ed. Cambridge, MA: The MIT Press; 1965.
42. Cárdenas AA, Amin S, Lin Z-S, Huang Y-L, Huang C-Y, Sastri S. Attacks against process control systems: risk assessment, detection, and response. In: *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*; March 22–24, 2011. Hong Kong, China:355. <https://doi.org/10.1145/1966913.1966959>.
43. Bennett S. The past of PID controllers. *IFAC Proc Vol*. 2000;33(4):1–11.
44. Gupta R, Chow MY. Networked control system: overview and research trends. *IEEE Trans Ind Electron*. 2010;57(7):2527–2535.
45. Ceccarelli A, Bondavalli A, Froemel B, Hoefberger O, Kopetz H. Basic concepts on system of systems. In: Bondavalli A, Bouchenak S, Kopetz H, eds. *Cyber-Physical Systems of Systems: Foundations - A Conceptual Model & Some Derivations: The AMAEOS Legacy*. Cham, Switzerland: Springer; 2016:257.
46. Xia F, Kong X, Xu Z. Cyber-physical control over wireless sensor and actuator networks with packet loss. In: Mazumder, S, ed. *Wireless Networking Based Control*. New York, NY: Springer; 2011:85–102.
47. Zhang D, Shi P, Wang QG, Yu L. Analysis and synthesis of networked control systems: a survey of recent advances and challenges. *ISA Trans*. 2017;66:376–392.
48. Yampolskiy M, Horváth P, Koutsoukos XD, Xue Y, Sztipanovits J. A language for describing attacks on cyber-physical systems. *Int J Crit Infrastruct Prot*. 2015;8:40–52.
49. Humayed A, Lin J, Li F, Luo B. Cyber-physical systems security - a survey. *IEEE Internet Things J*. 2017;4(6):1802–1831.
50. Khaitan SK, McCalley JD. Design techniques and applications of cyberphysical systems: a survey. *IEEE Syst J*. 2015;9(2):350–365.
51. Slay J, Miller M. Lessons learned from the Maroochy water breach. *IFIP International Federation for Information Processing*. 2007;253:73–82.
52. Amin S, Schwartz GA, Hussain A. In quest of benchmarking security risks to cyber-physical systems. *IEEE Netw*. 2013;27(1):19–24.
53. Hahn A, Thomas RK, Lozano I, Cardenas A. A multi-layered and kill-chain based security analysis framework for cyber-physical systems. *Int J Crit Infrastruct Prot*. 2015;11:39–50.
54. Peng Y, Lu T, Liu J, Gao Y, Guo X, Xie F. Cyber-physical system risk assessment. In: *Proceedings - 2013 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2013*; 2013. <https://doi.org/10.1109/IIH-MSP.2013.116>
55. Zhu Q, Rieger C, Başar T. A hierarchical security architecture for cyber-physical systems. *4th International Symposium on Resilient Control Systems*. Boise, ID: IEEE; 2011:15–20. <https://doi.org/10.1109/ISRCs.2011.6016081>.
56. Lee RM, Assante MJ, Conway T. German steel mill cyber attack. *Industrial Control Systems*; 2014:1–15. [http://ics3.sans.org/media/ICS-CPPE-case-Study-2-German-Steelworks\\_Facility.pdf](http://ics3.sans.org/media/ICS-CPPE-case-Study-2-German-Steelworks_Facility.pdf). Accessed February 22, 2019.
57. Perlroth N, Krauss C. A cyberattack in Saudi Arabia had a deadly goal. Experts fear another try. *The New York Times*. <https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html>. Accessed December 10, 2018.
58. Johnson B, Caban D, Krotofil M, Scali D, Brubaker N, Glycer C. Attackers deploy new ics attack framework “TRITON” and cause operational disruption to critical infrastructure. *FireEye*. <https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html>. Accessed April 12, 2018.
59. Pinto ADi, Dragoni Y, Carcano A. TRITON: the first ICS cyber attack on safety instrument systems. *Nozomi Networks*. <https://www.nozominetworks.com/downloads/US/Nozomi-Networks-TRITON-The-First-SIS-Cyberattack.pdf>. Accessed December 13, 2018.
60. Miller S, Reese E. A totally tubular treatise on TRITON and TriStation. *FireEye*. <https://www.fireeye.com/blog/threat-research/2018/06/totally-tubular-treatise-on-triton-and-tristation.html>. Accessed December 13, 2018.
61. Koscher K, Czeskis A, Roesner F, et al. Experimental security analysis of a modern automobile. *2010 IEEE Symposium on Security and Privacy*. Berkeley, CA: IEEE; 2010:447–462.
62. Drozhzhin A. Black hat USA 2015: the full story of how that Jeep was hacked. *Kaspersky Lab Daily*. <https://www.kaspersky.com/blog/blackhat-jeep-cherokee-hack-explained/9493/>. Accessed December 10, 2018.
63. DNV GL. Cyber security resilience management for ships and mobile offshore units in operation. *DNVGL-RP-0496*. 2016. <http://www.gard.no/Content/21865536/DNVGL-RP-0496.pdf>. Accessed October 30, 2018.
64. Torkildsen EN. Empirical Studies of Safety and Security Co-analysis of Autonomous Systems. [Master's thesis]. Trondheim: Norwegian University of Science and Technology. 2018.
65. Yampolskiy M, Horváth P, Koutsoukos XD, Xue Y, Sztipanovits J. Systematic analysis of cyber-attacks on CPS-evaluating applicability of DFD-based approach. *2012 5th International Symposium on Resilient Control Systems*. Salt Lake City, UT: IEEE; 2012:55–62.
66. Plioutsias A, Karanikas N, Chatzimihailidou MM. Hazard analysis and safety requirements for small drone operations: to what extent

- do popular drones embed safety?. *Risk Anal.* 2018;38(3):562-584. <https://doi.org/10.1111/risa.12867>.
67. Giraldo J, Sarkar E, Cardenas A, Maniatakos M, Kantarcioglu M. Security and privacy in cyber-physical systems: a survey of surveys. *IEEE Des Test.* 2017;7-17.
  68. Paul S, Brunel J, Rioux L, et al. *Recommendations for Security and Safety Co-Engineering (Release No. 3)*. MeRgE ITEA2 Project. 2016.
  69. Sun M, Mohan S, Sha L, Gunter C. Addressing safety and security contradictions in cyber-physical systems. In: *Proceedings of the 1st Workshop on Future Directions in Cyber-Physical Systems Security (CPSSW'09)*. Newark: US Department of Homeland Security. 2009.
  70. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Trans Syst Man, Cybern Part A Systems Humans.* 2000;30(3):286-297. <https://doi.org/10.1109/3468.844354>.
  71. SAE. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE Int.* 2016;J3016:1-12.
  72. Press kit: Metro automation facts, figures and trends. International Association of Public Transport. <https://www.uitp.org/metro-automation-facts-figures-and-trends>. Accessed November 22, 2018.
  73. Cyber-enabled ships: ShipRight procedure – autonomous ships. Lloyd's Register. <http://info.lr.org/l/12702/2016-07-07/32rrbk>. Accessed November 22, 2018.
  74. NBAA Automated Flight Deck Training Guidelines. National Business Aviation Association. <https://nbaa.org/press-releases/nbaa-publishes-automated-flight-deck-training-guidelines/>. Accessed November 22, 2018.
  75. Clough BT. Metrics, schmetrics! how the heck do you determine A UAV's autonomy anyway?. In: *Proceedings of the Performance Metrics for Intelligent Systems Workshop*. Gaithersburg, Maryland; 2002.
  76. Liu Y, Peng Y, Wang B, Yao S, Liu Z. Review on cyber-physical systems. *IEEE/CAA J Autom Sin.* 2017;4(1):27-40.
  77. Fridman L. Self-driving cars: state of the art (2019). <https://deeplearning.mit.edu>. Accessed February 5, 2019.
  78. Cardenas A, Amin S, Sastry S. Secure Control: Towards Survivable Cyber-Physical Systems. In: *The 28th International Conference on Distributed Computing Systems Workshops, Beijing, 2008*; 2008; 495-500.
  79. US Department of Homeland Security. Cyber physical systems security. <https://www.dhs.gov/science-and-technology/csd-cpssec>. Accessed May 10, 2018.
  80. Rasmussen J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans Syst Man Cybern.* 1983;13(3):257-266.
  81. Stanton NA, Stewart R, Harris D, et al. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics.* 2006;49(12-13):1288-1311.
  82. Rasmussen J. Human errors. A taxonomy for describing human malfunction in industrial installations. *J Occup Accid.* 1982;4:311-333.
  83. Endsley MR. Automation and situation awareness. In Parasuraman R, Mouloua M, eds, *Human Factors in Transportation. Automation and Human Performance: Theory and Applications*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1996:163-181.
  84. Taylor JR. *Human Error in Process Plant Design and Operations*. Boca Raton, FL: CRC Press; 2016.
  85. Orojloo H, Azgomi MA. A method for evaluating the consequence propagation of security attacks in cyber-physical systems. *Futur Gener Comput Syst.* 2017;67:57-71. <https://doi.org/10.1016/j.future.2016.07.016>.
  86. Magone A, Mazali T. *Industria 4.0. Uomini e Macchine Nella Fabbrica Digitale*. 1st ed. Milano: Guerini e Associati; 2016.
  87. Bever L. Amazon warehouse workers sickened after bear-spray incident. ABC News. <https://abcnews.go.com/WNT/video/amazon-warehouse-workers-sickened-bear-spray-incident-59637342>. Accessed December 10, 2018.
  88. Blue VJ. This wasn't even Amazon's first bear repellent accident. Wired. <https://www.wired.com/story/amazon-first-bear-repellent-accident/>. Accessed December 10, 2018.
  89. Boudette NE. Fatal Tesla crash raises new questions about autopilot system. New York Times. <https://www.nytimes.com/2018/03/31/business/tesla-crash-autopilot-musk.html>. Accessed December 10, 2018.
  90. Arnold RD, Wade JP. A definition of systems thinking: a systems approach. *Proc Comput Sci.* 2015;44(C):669-678.
  91. IEC. 61508-4 - Functional safety of electrical/electronic/programmable electronic safety-related systems - part 4: definitions and abbreviations. 2010.
  92. Herbert S. The architecture of complexity. In: *Proceedings of the American Philosophical Society*. Vol 106. American Philosophical Society; 1962:467-482.
  93. Hettinger LJ, Kirlik A, Goh YM, Buckle P. Modelling and simulation of complex sociotechnical systems: envisioning and analysing work environments. *Ergonomics.* 2015;58(4):600-614.
  94. Young W, Leveson NG. Systems thinking for safety and security. *Proc 2013 Annu Comput Secur Appl Conf*; 2013:1-8. <https://doi.org/10.1145/2523649.2530277>.
  95. Rausand M. *Reliability of Safety-Critical Systems. Theory and Applications*. Hoboken, NJ: John Wiley & Sons; 2014.
  96. Gilsinn JD, Schierholz R. Security assurance levels: a vector approach to describing security requirements. USDHS Industrial Control Systems Joint Working Group (ICSJWG). 2010. [http://www.nist.gov/customcf/get\\_pdf.cfm?pub\\_id=906330](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=906330). Accessed October 4, 2018.
  97. Rasmussen B, Grønberg CD. Accidents and risk control. *J Loss Prev Process Ind.* 1997;10(5-6):325-332.
  98. Dzung D, Naedele M, Von Hoff TP, Crevatin M. Security for industrial communication systems. *Proc IEEE.* 2005;93(6):1152-1177.
  99. National Cyber Security Centre, National Crime Agency. The cyber threat to UK business: 2017-2018 Report. National Cyber Security Centre National Crime Agency. <https://www.ncsc.gov.uk/cyberthreat>. Accessed November 21, 2018.
  100. He H, Maple C, Watson T, et al. The security challenges in the IoT enabled cyber-physical systems and opportunities for evolutionary computing & other computational intelligence. In: *2016 IEEE Congress on Evolutionary Computation (CEC)*. 2016. <https://doi.org/10.1109/CEC.2016.7743900>.
  101. Rinaldi SM. Modeling and simulating critical infrastructures and their interdependencies. *IEEE Control Syst Mag.* 2001;21(6):11-25.
  102. Ouyang M. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliab Eng Syst Saf.* 2014;121:43-60. <https://doi.org/10.1016/j.res.2013.06.040>.
  103. Hagen IB, Kufoalor DKM, Brekke EF, Johansen TA. MPC-based collision avoidance strategy for existing marine vessel guidance systems. *Proc IEEE Int Conf Robot Autom*; 2018:7618-7623.
  104. Carreras Guzman NH, Kozine I. Uncontrolled flows of information and energy in cyber-physical systems. European Safety and Reliability Association Newsletter. 2018. <http://www.esrahomepage.eu/filehandler.ashx?file=16438>. Accessed November 19, 2018.

## AUTHOR BIOGRAPHIES



NELSON H. CARRERAS GUZMAN is a PhD Fellow in the Engineering Systems Group at the Technical University of Denmark (DTU). He earned a BSc. degree in Civil Engineering from the Catholic University Andres Bello in Caracas, Venezuela, and a MSc. degree in Civil Engineering for Risk Mitigation from Politecnico Di Milano, Italy. As a civil engineer, he worked in hydraulic engineering design and hydrological modeling. Nelson is experienced in Critical Infrastructure Protection and Resilience projects, where he has worked in collaboration with Civil Protection authorities, infrastructure operators, and research agencies. His main research interest is the integration of the safety and security fields, enhancing risk science to support responsible innovation in cyber-physical systems and critical infrastructures.



MORTEN WIED, cand.techn.soc., currently holds a position as PhD Fellow at the Technical University of Denmark and Associated Senior Consultant at the private consultancy Let's Involve. Morten functions as advisor and analyst for businesses, foundations, and universities involved in large technology development projects. Over the years, he has worked in the fields of energy, transport, health, agriculture, and aerospace. Morten's research interests lie in the overlap between decision science, systems theory and project management. Morten has worked in private consultancy since 2009, prior to which he held a position as Head of Section at the Danish Ministry of Science, Technology, and Innovation. Prior still, he worked as a research assistant with Risø National Laboratories as part of the Research Programme for Technology Scenarios. He holds a master's degree in Technology and Socioeconomic Planning (cand.techn.soc.) from the University of Roskilde.



IGOR KOZINE has been a senior researcher at the Technical University of Denmark, Department of Management engineering since 2009. Before that, he had worked as a senior researcher at Risø National Laboratory, Denmark, for 10 years. The earlier positions as a trainee-scientist, researcher and associate professor he had at the Institute of Nuclear Power Engineering, Obninsk, Russia. For one year, he did his research study as a Fulbright

Scholar in the State University of New York at Binghamton. He studied at the Moscow Institute of Physics and Engineering (Technical University) and received his M.S. and Ph.D. in Systems Sciences from the same university. His research area is reliability and safety analysis of safety-critical systems. The focus of particular interest has been on uncertainty representation in risk and reliability models, simulation of human performance and resilience assessment of critical infrastructure.



MARY ANN LUNDTTEIGEN has been a professor in Department of Mechanical and Industrial Engineering since 2011, with a period with DNV-GL as Principle Engineer from 2012–2013. She has a PhD in reliability of safety-instrumented systems (2009), and a MSc. in engineering cybernetics (1993). Before starting on her PhD, she worked for several years in industry, including as instrumentation engineer (onshore) and automation and electrical supervisor (offshore) in Phillips Petroleum, automation leader at the factory of Nidar, and senior researcher at SINTEF, department for applied cybernetics. Her main research focus and interest concern functional safety and reliability of safety-related electrical/electronic/programmable electronic (E/E/PE) systems. She is also a member of IEC 61511 committee who maintains the standard on functional safety for process industry sector. As a co-director and responsible for reliability subject in the 8-year research center on subsea production and processing (SUBPRO, see <http://www.ntnu.edu/subpro>), she has extended her research to cover reliability, safety, and condition-based maintenance of systems with particular demanding environmental and operational conditions, such as subsea systems. Lundteigen has a long-lasting and extensive contact network in Norwegian industry work, due to involvement in SINTEF projects over several years, many of them through the PDS forum (<http://www.sintef.edu/pds>). She has more than 50 publications peer-reviewed papers, including around 20 papers in international journals. She has also been contributing to a high number of studies for industry companies.

**How to cite this article:** Carreras Guzman NH, Wied M, Kozine I, Lundteigen MA. Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis. *Systems Engineering*. 2020;23:189–210. <https://doi.org/10.1002/sys.21509>

# APPENDIX A: Detailed CPS master diagram of autonomous surface vehicle, highlighting hazardous energy flows and a selection of types of attacks

