Intelligent Security Measures for Smart Cyber-Physical Systems

Muhammad Shafique*, Faiq Khalid*, Semeen Rehman†
*Institute of Computer Engineering, Vienna University of Technology (TU Wien), Austria
†Institute of Computer Technology, Vienna University of Technology (TU Wien), Austria
{muhammad.shafique; faiq.khalid; semeen.rehman}@tuwien.ac.at

***Abstract*— The exponential growth of cyber-physical systems (CPS), especially in safety-critical applications, has imposed several security threats (like manipulation of communication channels, hardware components, and associated software) due to complex cybernetics and the interaction among (independent) CPS domains. These security threats have led to the development of different static as well as adaptive detection and protection techniques on different layers of the CPS stack, e.g., cross-layer and intra-layer connectivity. This paper first presents a brief overview of various security threats at different CPS layers, their respective threat models and associated research challenges to develop robust security measures. Moreover, this paper provides a brief yet comprehensive survey of the state-of-the-art static and adaptive techniques for detection and prevention, and their inherent limitations, i.e., incapability to capture the dormant or uncertainty-based runtime security attacks. To address these challenges, this paper also discusses the intelligent security measures (using machine learning-based techniques) against several characterized attacks on different layers of the CPS stack. Furthermore, we identify the associated challenges and open research problems in developing intelligent security measures for CPS. Towards the end, we provide an overview of our project on security for smart CPS along with important analyses.**

Keywords*— *Cyber-Physical Systems, CPS, Machine Learning, Neural Networks, Deep Learning, DNNs, Security, Attacks, Static and Dynamic Techniques, Attack Surface, Autonomous Vehicle, Intelligent Measures.

I. INTRODUCTION

The rapid advancements in communication, computing, and physical control systems have enabled the cyber world to integrate and closely interact with the physical domain that leads to Cyber-Physical Systems (CPS) [1]. CPS is the integration of embedded computing devices, smart objects, people and physical environments, which are tightly coupled via communication networks [2]. Typically, CPS collects the information from the physical domain and analyzes it to issue the appropriate control commands, as shown in Fig. 1. Due to their ability to control the state of the systems, with respect to the physical characteristics, these systems are being widely used in several safety critical domains, e.g., intelligent traffic control and infrastructure management (Smart Cities), healthcare, transport systems, industrial automation (Smart Factories), power distribution and generation (Smart Grids), autonomous vehicles (Smart Cars) and automated houses (Smart Buildings, Smart Homes) [3][4]. Fig. 1 shows the different physical (i.e., acceleration, gearbox, steering, airbag, light, braking, charging and tire measurement controls) and cyber components (i.e., vehicle-to-vehicle communication, GPS tracking/navigation, Wi-Fi, Bluetooth connectivity of multiple devices) in smart autonomous vehicles.

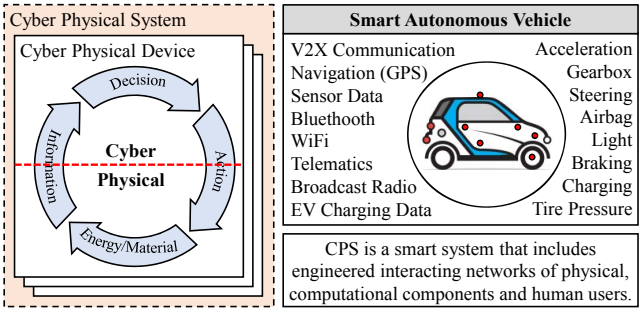

Fig. 1. Key features of a CPS and for an example of Smart Autonomous Vehicle. Red-line in this figure separates the Cyber and physical worlds.

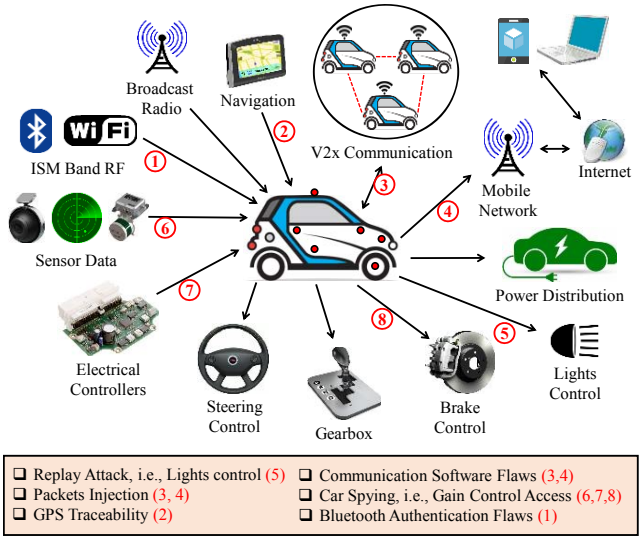

Fig. 2. Examples of Security Threats in Smart Autonomous Vehicles.

The massive integration of cyber domain (i.e., networked computing devices), physical domain (i.e., actuators) and humans are playing a significant role in the rise of the *Internet-of-CPS-Things*, where different cyber-physical sub-systems are integrated to realize high-end smart services [1]. However, these systems are becoming more and more vulnerable to various security threats at different layers of the hardware and software CPS stacks, covering both computation and communication layers. Consequently, several security incidents related to physical and hardware attacks on the cyber-physical systems have been reported in real-world. Some of the most prominent incidents are city water pipeline [5], pacemaker [6], ABS wheel

978-1-5386-7377-5/18/$31.00 ©2018 IEEE
DOI 10.1109/DSD.2018.00058

speed sensor spoofing in smart cars [19] and several industrial attacks [7][8][9]. Fig. 2 shows some of the possible cyber (i.e., packet injection, GPS traceability, communication software and Bluetooth authentication flaws) and physical attacks (gain control attacks) in smart autonomous vehicles. At the same time, security features need to be adaptive yet energy-efficient to account for unpredictable operational scenarios, even years after the manufacturing and deployment of an autonomous vehicle will stay in the field for several decades. Therefore, such systems have to meet stringent design requirements in terms of security, adaptability [10], dependability [11], and energy efficiency [3]. Moreover, the security features need to be intelligent to combat with various attack models, which could even be unforeseen at the design time.

This paper makes the following novel contributions:

1) A *brief yet comprehensive overview* of the CPS security including various security threats at different CPS and their respective threat models.
2) Highlights of the *associated research challenge* to develop robust security measures for security threats at different CPS layers.
3) A *brief survey of the state-of-the-art* security measures along with a discussion on their pros and cons.
4) An overview of our project on intelligent security measures for smart CPS (Sec4SCPS) along with important analyses.

Paper Organization: Section II provides a brief overview of inter-/intra-layer security threats and the respective threat models. Section III highlights the design challenges for developing secure smart CPS. Section IV discusses and identifies the key limitations of the traditional, adaptive and intelligent security measure for smart CPS. V provides a brief overview of our on-going project on the intelligent security measures for smart CPS, following by the conclusion in Section VI.

II. SECURITY FOR CPS

To provide the better understanding of the security measures for CPS, in this section, we provide a brief overview of several security attacks/vulnerabilities with respect to CPS layers, their respective payloads, and associated threat models.

A. Security Threats in CPS

Unlike the traditional systems, the security threats in CPS are also dependent on the uncertain behavior of physical domain which led to several safety and security critical scenarios in the physical domain [4][12]. Therefore, it is required to categorize the security threats for effective security measures. Based on the CPS layers, the security attacks can be categorized as follow (see: Table 2, that summarize some of the possible and real-world attacks on each CPS layer [13]):

1) In the **physical layer,** an attacker can launch direct intervention or destruction the physical objects to monitor and control, the sensors and controller which results in inaccurate sensed measurements [14], incorrect control decisions, and inappropriate actuator actions, as shown in Fig. 3. Table Table 1 provides some examples of physical attacks in smart grids, smart cars, and smart healthcare.

TABLE 1. EXAMPLE OF THE PHYSICAL ATTACKS ON SMART GRIDS, SMART HEALTHCARE, AND SMART CARS

Application	Attack	Description
Smart Grids	Natural Threats	In 2014, wild animals cause 150 blackouts in the US [18].
	Accidental Threats	In US, cars hitting the transmission line cause the 356 blackouts in 2014 [18].
	Material theft	Copper wire theft caused the blackout [18].
Smart Healthcare	Acquiring unique IDs	Obtaining devices IDs [19]
Smart Cars	ABS wheel speed sensor spoofing	Disruption of the magnetic field around the sensor introduces the incorrect measurements [19].

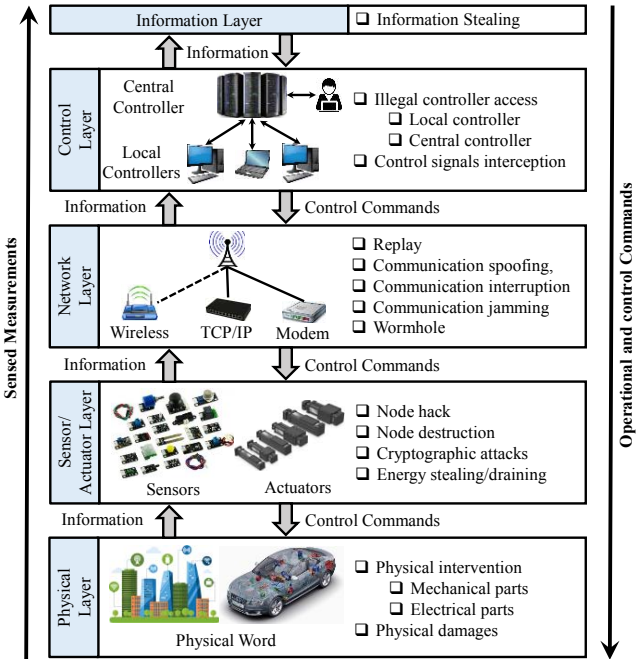

Fig. 3. CPS Architecture and Intra Layer Security Threats

2) The security of **sensor/actuators networks** is one of the key components of CPS, which is mostly ignored in CPS security [15]. However, the confidentiality and sensitivity of the sensed/measured data make it one of the most vulnerable layers with respect to the security threats (i.e., privacy) [16]. There are the following possible ways to attack this layer:
   a. An attacker can *destroy/hack* the sensors/actuators with brute force attacks to extract the sensitive information from the sensors/actuators (e.g., secret keys, side channel parameters), and can also *modify/manipulate* them [17].
   b. An attacker can hack the power distribution mechanisms of sensor/actuators to *drain the energy* for denial-of-service attacks or to use that energy to activate the malicious circuitry/payloads [17].

TABLE 2. SECURITY THREATS AT DIFFERENT CPS LAYERS [13]

Layers	Attacks	Details	Attack type
Physical Layer	Direct intervention and damages	Changes the hardware or mechanical parts to damage the system.	Denial-of-service, half-life reduction
Sensor/Actuator Layer	Node hacking	Leakages the information directly from sensors/actuator via RF signals.	Information leakage
	Node destruction	Destructs, extracts, or modifies node physically.	Denial-of-service
	Energy stealing/draining	Quickly drains out the limited power of sensors/actuators.	Denial-of-service,
	Cryptographic attacks	Cracks secret keys with brute force, dictionary, monitoring, or side channel analysis.	Information leakage
Network Layer	Replay	Forwards message to an incorrect destination or with a delay.	Timing attacks, information leakage
	Communication Jamming	Halts the on-going communication.	Denial-of-service
	Data Flooding in communication	Inserts the bogus data into the established communication to overload the system.	Denial-of-service
	Sybil	An adversary illegitimately takes on multiple identities	Denial-of-service, information leakage
	Spoofing and altering the communication information	Changes routing information illegitimately	Denial-of-service, information leakage
	Wormhole	Disrupts the routing	Denial-of-service
	Selective communication	Disrupts the on-going communication and sends only selective data	Denial-of-service, information leakage
Control Layer	Controller hacking	Hacks the controller to perform malicious activities	Denial-of-service, information leakage, timing attacks
	Control signal hacking	Interrupts and manipulate the control signals to perform malicious activities	Denial-of-service, information leakage, timing attacks
Information Layer	Privacy	Steal information from eavesdropping and traffic analysis	Information leakage

a. In case of the sensors/actuators-based security keys, an attacker can *extract the key* using brute force, dictionary attack or monitoring attack [17].

3) Typically, the security threats in the **network layer** of CPS are related to the communications [19]. There are two major types of networking attacks; 1) *Replay Attack*: in this attack, the message is forwarded to an incorrect destination, or to the destination with a delay [22]. 2) *Denial of service (DoS)*: in this attack, a malicious event can diminish or eliminate the network capacity to perform its expected functions [23]. DoS can further be divided into following attacks:

a. *Jamming:* In this attack, an attacker can jam a node or a group of nodes by signal interference [24].

b. *Collision:* In this attack, an attacker can force the system to violate the communication protocols and continually transmit messages to generate inconsistencies [23].

c. *Routing ill-directing:* In this attack, an attacker can force the system to refuse the route messages in terms of the nature of multi-hop communications[23].

d. *Flooding:* In this attack, an attacker can send unnecessary connection requests to a vulnerable node [23].

e. *Wormhole:* In this attack, an attacker can create a well-placed wormhole for disrupting the network routing [23].

f. Selective forwarding: In this attack, an attacker can force the intruded node to forward the messages selectively to disrupt the data transmission. For example, *Sybil attack* [25] is done by such adversary that illegitimately takes multiple identities to reduce the availability of a network by spoofing, altering, and replaying routing information [23].

4) Typical security threats in the **control layer** of CPS are from desynchronization [26] because control mechanisms are highly dependent on timeliness. Therefore, a slight desynchronization in the control units/signals can be considered as catastrophic because of the sudden incorrect decisions can lead to CPS failures.

5) At the information layer, most of the attacks perform information stealing by either eavesdropping or analyzing the traffic data. However, the manipulation [79] of the key information/data can also be used to perform other attacks, i.e., jamming, collision, denial of services, etc.

B. CPS Security Threat Models

To develop the security measures against any security vulnerabilities, the first step is to define the threat model for getting the better understanding of the attack strength and attacker capabilities. In CPS, an attack is defined as *a sequence of events that force the CPS to deviate from it's anticipated or specified execution flow, with the intention of breaching one or more security objectives* [27], i.e., confidentiality, integrity and authenticity of control commands and availability of the CPS. Therefore, to define a certain threat model, the following factors have to be identified for a particular attack:

1) *Source/Attacker:* It is defined as anything which can disturb or interrupt the behavior or functionality of the CPS [27]. It is not necessary that an attacker can only be an organization, individual or state/nation because all the accidental events and environmental disasters can also be considered as a source of an attack.

2) *Target:* It is defined as the targeted layer or device that a source is trying to get the access [27].

3) *Motive:* It is defined as the reasons to launch an attack, i.e., criminal, spying, terroristic, political, or cyberwar [27].

4) *Attack Vector:* It is defined as the mechanisms that can be used to perform successful attacks, i.e., interception, interruption, modification, and fabrication [27].

5) *Payload:* It is defined the consequences of the successful attacks, i.e., confidentiality, integrity, availability, privacy, or safety [27]. Some of the possible payloads related above-mentioned security attacks are following:

   a) *Availability:* In this payload, an attacker can get the access to the control units/signals for making the system/data unavailable for each other [2][28][29].

   b) *Timing Constraint:* In this payload, an attacker can interrupt the computations/executions of the task to miss the completion deadline [30].

   c) *Eavesdropping:* In this payload, an attacker observes the CPS operations without any interference [28].

   d) *Compromised-Key:* In this payload, an attacker can get a secured communication without the perception of sender or receiver by using the compromised key [29].

   e) *Denial-of-Service:* In this payload, an attacker can block the communication (by overloading the traffic) or halts system operations.

III. Design Challenges for Security Measures of Cyber-Physical System

The security threats and vulnerabilities discussed in Section II.A have raised the following key research challenges to design the secure CPS:

1) **Security by Design:** Due to exponential growth in usage of CPS and faster time to market, the security threats are not being considered as the fundamental design challenges in CPS design cycle [31]. Though several security measures are being introduced, most of them are focused towards the cyber-attacks. However, the intentional (malicious intent) or unintentional (natural or environmental disaster) physical attacks pose several critical design concerns. In results, a key question arises that *how to embed the physical security measures during the CPS design cycle?*

2) Real-timeliness Nature: Typically, the CPS are being tested and analyzed for the security vulnerabilities during the design stage or before the deployment stage (post-fabrication testing stage). However, several unforeseen (natural, environmental or accidental disasters) and intentional threats can occur during the runtime [32]. In order to apply certain security measures during the runtime, it is crucial to introduce the runtime detection and decision capability in the CPS. However, due to resource constraints and energy budget, it limits the scope of runtime security measures which raises a fundamental research question that *how to design a certain runtime security measure with maximum coverage of security vulnerabilities while considering the resource constraints and energy budget?*

3) **Secure Integration:** Typically, the CPSs are very complex and involve several stakeholders, especially, in the integration of heterogeneous cyber-physical devices. This heterogeneity makes it very challenging to integrate such variety of cyber-physical components in a secure way. Therefore, a research question arises that *how to ensure the secure integration of several heterogeneous cyber-physical devices to develop the secure CPS?*

4) **Privacy:** Most of the CPSs measured or sensed confidential information through different sensors and communicate this data to several devices via multiple communication channels which leads to a very critical security challenge, i.e., privacy. Therefore, a key research question arises that *how to ensure privacy the information while measuring it from sensors and securely communicate it to other cyber-physical devices?*

IV. Security Measures of CPS

To address the design challenges that are mentioned in Section III, for developing the secure CPS, several security measures have been proposed. Depending upon the methodologies, we have categorized them into following two categories:

A. Traditional Security Measures

Several security measures [33][34] have been proposed to prevent cyber-physical devices from the security threats discussed in Section II. For example, Physical Unclonable Functions (PUFs) [35][36][37][38] and True Random Number Generator (TRNGs) [39][40] based prevention and anti-tampering techniques have been used for secure communication and interaction with the physical world. One of the key advantages of these techniques is that they can generate the necessary keys and authentication IDs, without requiring any on-device key storage mechanism, and can provide the obfuscation against tempering and reverse engineering. Moreover, these security measures center around random faults [41] during design or fabrication stages and not the ones that can be stealthy during the testing stage or even at the earlier stages of runtime operations. However, these techniques are based on design time solution and can only be applicable at testing stage (before deployment), therefore, these techniques have the following limitations:

1) Unable to *incorporate the effects of stealthy attack*s during the testing stage or runtime.

2) Unable to *incorporate the effects of uncertainties* during the real-world scenarios.

B. Adaptive Security Measures

In order to address the above-mentioned limitations of the traditional security measures, several runtime techniques have been proposed [42][43][44] which can adapt to incorporate uncertainties and stealthy attacks. Since the CPS security can be breached at any stage of its design cycle or workflow, i.e., designing, fabricating, testing or runtime, and sensing, computing, communicating or actuating [44]. Therefore, several sensors [44] and context-dependent [46][47] security measures have been proposed. Similarly, industrial CPS, such as smart grid and critical infrastructures' security Is based on the idea of ensuring security from the control systems perspective [49][50].

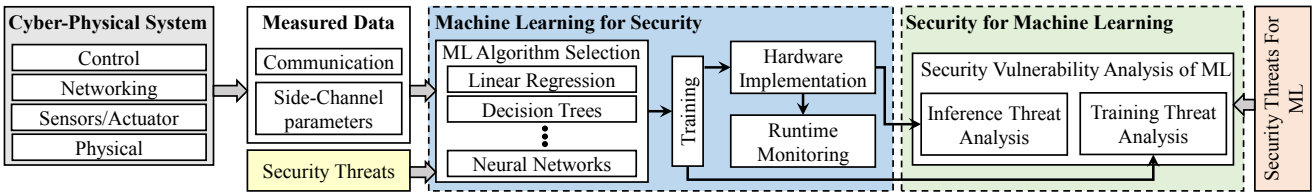

Fig. 4. A brief overview of our project Security for Smart CPS (Sec4SCPS).

[50][7][51] and at a roundtable discussion [52]. It was suggested in [52] to work on a language, or a feature of it, to let the designers work on the security enhancement in synchronization with other requirements of CPS. Indeed, there has been little effort, notably from the control theory perspective [53], in considering security as a design parameter for CPS from an early design phase. Though these security measures provide the comprehensive runtime solutions for securing the CPS due to exponential growth in number cyber-physical devices, these *traditional adaptive security measures are not sufficient to incorporate the runtime computational needs for security measures*.

C. Intelligent Security Measures for CPS

To address the above-mentioned challenges of the adaptive security measures, several machine learning (ML) [54][55][56][57][58][59], specifically neural networks based approaches have been proposed because of the ability to extract the hidden features form the big amount of sensed/measured data [60]. Fig. 5. shows design flow of the ML-based security measures for CPSs which consists of the following steps:

1) *System Modeling:* Due to heterogeneity in cyber-physical devices, it is nearly impossible to build each device in a secure environment which makes all the measurements from the devices unusable for training the machine learning algorithm. Therefore, the first step is model the device abstract behavior, depending upon its complexity, to generate the data for training the machine learning algorithms, as shown in Fig. 5.

2) *Security Analysis* with respect to selected trained ML algorithm: in the next step, depending upon the security parameters, design constraints, i.e., power and area, and the complexity of the generated data, an appropriate machine learning algorithm is selected and then trained for the acquired data. Then this trained ML algorithm is used to analyze and detect the different security aspects and anomalous behavior in CPSs, i.e., power behavior analysis for computing cores.

Though several works have been proposed based on the above-mentioned methodology most of them are focused on encryption block related challenges, i.e., information leakage. However, this methodology can be extended to analyze other side channel parameters [61][62][63] and communication patterns [64][65]. Therefore, we are currently working on the project (Smart security for CPS (Smart Sec4CPS)) which exploits of side channel parameters, i.e., power, and communication behavior to detect the security attacks during runtime, which is briefly explained in the next section.

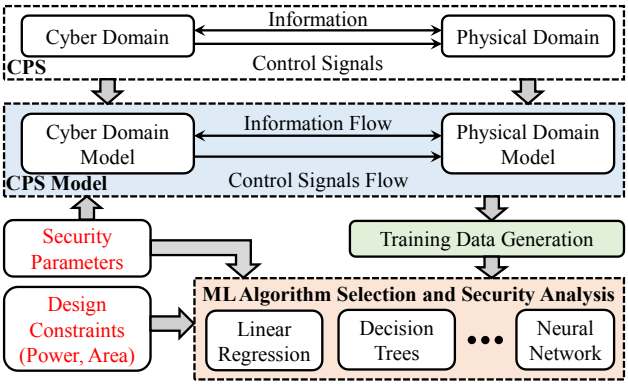

Fig. 5. Framework for Adaptive and Intelligent Security Measures.

V. SEC4SCPS

In this section, we provide an overview of our on-going project related to intelligent security measures for smart CPS (Sec4SCPS). In this project, we actively investigating different techniques for Hardware Security and Machine Learning Security to develop the secure CPS. Fig. 4 provides a brief overview of Smart Sec4CPS, which shows that based on research challenges this project has two major following research areas:

A. Machine Learning for Security (ML4Sec)

To address the issue of big data analysis for CPS security, in this project, we explore several parametric [61][62][63] and communication behaviors [64][65] (i.e., power and communication behavior) to improve the effectiveness of ML-based security measures. In order to choose the suitable parameters, first, we analyze the power behavior of the MC8051 with and without available intrusion benchmark (trust-hub [66]), i.e., MC8051-T200, as shown in Fig. 6. It illustrates that the power distribution with respect to pipeline stages is dependent on the instructions. Moreover, it shows that an intrusion (MC8051-T200) has the significant impact on the power distribution, as depicted from the comparison between label 1 and 3, and label 2 and 4. In result, we conclude that the power behavior with respect to pipeline stages can be used to identify the abnormalities in computing core, e.g., MC8051. However, power behavior modeling and measurement during runtime is not easy and poses the following research challenges:

1) How to *model the power behavior* in such a way that it can be used during runtime measurements?

2) How to reduce the measurement (power-ports) and runtime modeling/ measurement *area and energy overhead?*

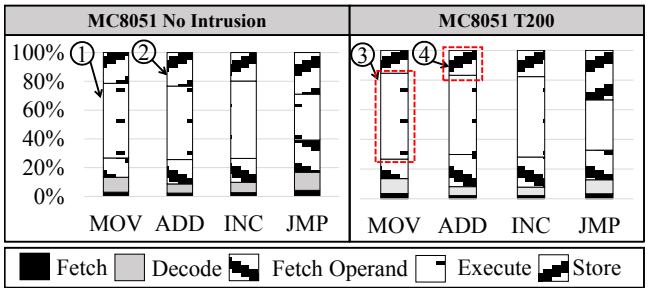

Fig. 6. Effects of Intrusions on Power Correlation with respect to Pipeline Stages for Different Instructions, i.e., MOV, ADD, INC, JMP.

Similarly, we analyzed the MC8051-T200 effects on UART communication for MC8051. The analysis in Fig. 7 shows that sometimes the output packets of the communication channels are less (in case of denial of service attacks) than the input ones and vice versa (in case of flooding, jamming, and information leakage attacks). Therefore, sophisticated analysis of the communication behavior of CPS devices can be used to find the abnormalities during runtime. However, this poses following research challenges:

1) How to *model the communication behavior* for runtime monitoring with minimum overhead?

2) How to *measure and analyze the communication* during runtime with minimal area and energy overhead?

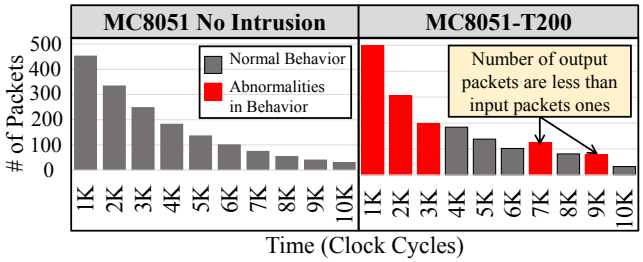

Fig. 7. Effects of Intrusion on Communication Behavior.

B. Security for Machine Learning (Sec4ML)

ML algorithms possess the inherent security vulnerabilities which can be manipulated to perform the security attacks[67]. In the design process of ML-based security measures, first, an ML algorithm is trained and validated based on the training dataset, and then the trained ML algorithm is used for inference, as shown in Fig. 8 (see label A and B: which refer the training and inference data poisoning). As the process to develop the ML-based techniques is dependent on several data dependencies and complex computations, which makes it vulnerable to several security attacks during training and inferencing stages, i.e., data poisoning during training and inference stages, and ML architectural intrusions. These attacks can be catastrophic for the performance, accuracy, and reliability of the deployed ML algorithm. Typically, based on the attacker's goal (payload of the security threats), the security attacks, for ML algorithms, can be divided into the following categories [68]:

1) *Confidence Reduction:* In this attack, an attacker can introduce the ambiguity in classification to reduce the confidence level (defined as the entropy of the output class probabilities) of output classes.

2) *Random Misclassification:* In this attack, an attacker can change the output classification to a random output class different from original class.

3) *Targeted Misclassification:* In this attack, an attacker can produce the inputs or can intrude the ML architecture that can force the output classification to a specific target class different from original class.

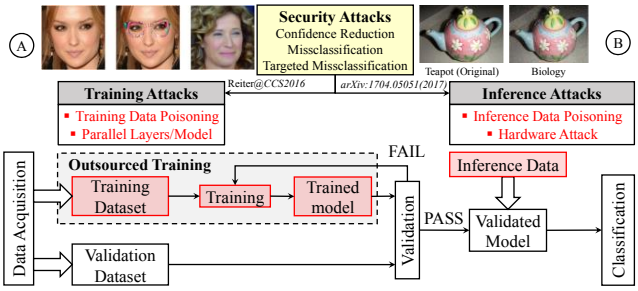

Fig. 8. Different Types of Security Attacks on Machine Learning Systems.

The strengths and weaknesses of the above-mentioned security attacks depend upon the attacker's capabilities to get the access for hardware implementation, tools and dataset, which in combination with attack types can be referred as an *attack surface*. For example, the attack surfaces shown in Fig. 9, shows the strength and difficulty level of security attacks during the training and inference stages, respectively. Therefore, the following subsections briefly discuss security vulnerabilities of ML algorithm during its training and inference stages and some state-of-the-art attacks.

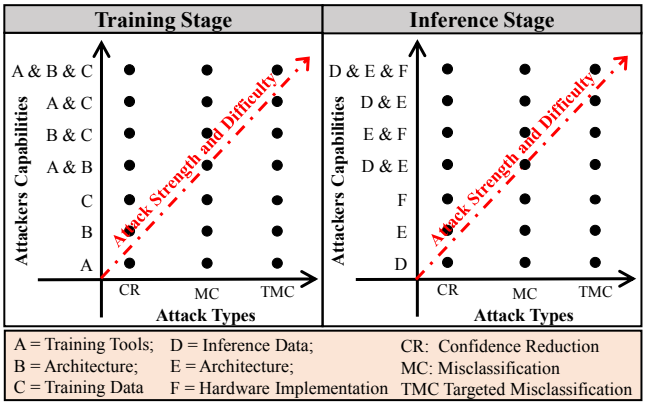

Fig. 9. Attack Surface for Machine Learning Algorithms during Training and Inference Stages with respect to Attacker Capabilities to access the datasets, tools and DNN architecture (including all hyperparameters) and Payload (the outcome of an attack, i.e., targeted or non-targeted misclassification) of the Security Attacks.

1) Security Vulnerabilities during Training Stage

In the training stage, the model parameters are obtained from the available training dataset which is assumed to capture the input space of the overall system. However, for larger datasets, it is not always feasible to locally train the ML algorithms, especially neural network, because of limited computational resources and high non-recurring engineering (NRE) cost [69]. Therefore, either outsourced (third-party cloud services providers) training or transfer learning is used, which increases the possibility of security attacks. The security attacks during

training are highly data dependent and their effectiveness is measured based on the attack payloads and attacker capabilities to get the illegal access to training datasets, tools, and the ML architectures. For instance, in the training part of the Fig. 9, the most powerful attack (the right top dot of the subgraph for training) is when the attacker has the access to all training steps, devices, and underlying algorithms, i.e., training tools, data and algorithm architecture. Thus, the attacker can train the model to alter the output classification of a specific input class to the target class (different from the original class).

Most of the state-of-the-art security attacks, during the training stage, are focusing on *training data poisoning* to launch different payloads of security attacks, e.g., confidence reduction [70], random [71] and targeted[71] [72][73] misclassifications. Similarly, other *architectural modification-based attacks* can also perform the targeted misclassifications [74]. Although the architectural modification and training data poisoning attacks are very effective they can affect the inference accuracy. Therefore, there are many alternatives which can be explored to generate an effective attack without reducing the inference accuracy. For example, training data poisoning with respect to architecture knowledge or manipulation training tools, Therefore, in outsourced training, several prevention techniques have been proposed and one of the most commonly used is encrypting the training dataset before outsourcing [75].

2) *Security Vulnerabilities during Inference Stage*

The inference stage in ML algorithms is also vulnerable and data dependent. However, the data poisoning attacks during the inference stage are not effective because a sophisticated preprocessing stage can significantly reduce intruded patterns in real-time data. Thus, inference data poisoning-based attacks are the weakest and easy to prevent at inference stages, as shown by the attack surface in Fig. 9 (see: inference subgraph). However, highly correlated patterns can be generated and intruded by combining the inference and inference data poisoning. Therefore, the possibility of such attacks, that are highly correlated and with high structural similarity index, cannot be ignored because preprocessing stage can overlook very high correlated intruded patterns. For example, Sharif et al. proposed an attack which introduces a glasses shape perturbation in training dataset and then it uses this pattern to launch a misclassification attack on neural network based face recognition system [72], as shown in Fig. 8 (see: label A). Alternatively, an attacker can exploit the data acquisition block [76], architecture or its hardware implementations [77]. Although preventions of such attacks are not easy because, in most of the cases, trained models run on third-party hardware which can have the multiple dormant or active intrusions [78].

Therefore, based on the above discussion on security vulnerabilities in ML algorithms during its training and inference stages, we identify the following research challenges to ensure security and privacy in machine learning based security measures.

1) How to *ensure the confidentiality and privacy* of the training datasets and its corresponding labels, especially for outsourced training and transfer learning?
2) How to the *ensure the secure and isolated* data acquisition during the inference stage?
3) How to *identify and prevent the highly correlated data* intrusion in dataset during the pre-processing phase.
4) How *to ensure the secure hardware implementation* to prevent the dormant and active intrusions in third-party hardware accelerators for ML algorithm?

VI. Conclusion

This paper presents a brief analysis of the security threats at different CPS layers and their respective threat models and identified the associated research challenges to develop secure CPS. To address these challenges, this paper also a comparative analysis of the state-of-the-art static and adaptive techniques for detection and prevention, and their associated limitations. In the end, this paper discusses the ML-based security techniques against several characterized attacks on different layers of the CPS and identifies open research problems in developing the intelligent security measures for CPS. Furthermore, the paper provides an overview of our on-going project related to security for CPS and discusses the research problems with corresponding motivational analyses.

Acknowledgment

This work is supported in parts by the Austrian Research Promotion Agency (FFG) and the Austrian Federal Ministry for Transport, Innovation, and Technology (BMVIT) under the "ICT of the Future" project, IoT4CPS: Trustworthy IoT for Cyber-Physical Systems.

References

[1] R. F, Babiceanu et al. "Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook", in Computers in Industry, 81,2016, pp. 128-137.
[2] A. Chattopadhyay et al. "Secure Cyber-Physical Systems: Current trends, tools, and open research problems", in IEEE **DATE,** 2017, pp. 132-145.
[3] S. Rehman et al. "Security Requirements Engineering (SRE) Framework for Cyber-Physical Systems (CPS): SRE for CPS", in **SoMeT**. Vol. 297. IOS Press, 2017.
[4] M. Wolf et al. "Safety and Security in Cyber-Physical Systems and Internet-of-Things Systems", in Proceedings of the IEEE, 106(1), 2018, pp. 9-20.
[5] Verizon, "Data Breach Digest", Issue Februray 2017. http://www.verizonenterprise.com/resources/reports/rp_data-breachdigest_xg_en.pdf.
[6] Marie Moe, "Go Ahead, Hackers. Break My Heart", (2017). https://www.wired.com/2016/03/go-ahead-hackers-break-heart/
[7] H. Song et al. "Security and Privacy in Cyber-physical Systems: Foundations, Principles, and Applications", in John Wiley & Sons, 2017.
[8] F. D. Priscoli et al. "Ensuring cyber-security in smart railway surveillance with SHIELD", in **IJCCS,** 2017,: pp. 138-170.
[9] D. Antonioli et al. "Taking Control: Design and Implementation of Botnets for Cyber-Physical Attacks with CPSBot", arXiv:1802.00152 (2018).
[10] H. Muccini et al. "Self-adaptation for cyber-physical systems: a systematic literature review", in ACM **SEAMS**, 2016.
[11] T. S. Alemayehu et al. "Dependability analysis of cyber-physical systems", in IET Computers & Digital Techniques, 11(6),2017, pp. 231-236.
[12] R. S. Chhetri et al. "Cross-domain security of cyber-physical systems", in **ASP-DAC**, 2017, pp. 200-215.
[13] S. Han et al. "Intrusion detection in cyber-physical systems: Techniques and challenges", in IEEE systems journal, 8(4), 2014, pp. 1052-1062.
[14] J. Wurm et al. "Introduction to cyber-physical system security: A cross-layer perspective" in IEEE **TMCS, 3**(3), 2017, pp. 215-227.
[15] A. Chattopadhyay et al. "Security of autonomous vehicle as a cyber-physical system", in IEEE, ISED 2017, pp. 1-6.
[16] J. A. Stankovic. "Research directions for the internet of things." IEEE Internet of Things Journal, 1(1), 2014,pp. 3-9.

[17] Y. Li et al. “Controllability and observability of CPSs under networked adversarial attacks.” IET Control Theory & Applications, 11(10), 2017, pp. 1596-1602.

[18] Powering Business Worldwide Eaton. (2014). Power Outage Annual Report: Blackout Tracker. [Online]. Available: http://www.eaton.com/blackouttracker

[19] J. Radcliffe, “Hacking medical devices for fun and insulin: Breaking the human SCADA system,” in Black Hat Conference Presentation Slides, 2011.

[20] Y. Shoukry et al. “Non-invasive spoofing attacks for anti-lock braking systems,” in Springer **CHES,** 2013, pp. 55-72.

[21] S. Ali et al. “WSN Security Mechanisms for CPS.” In Cyber Security for Cyber-Physical Systems, Springer, Cham, pp. 65-87.

[22] Y. Mo et al. “Secure control against replay attacks.” Allerton Conference on Communication, Control, and Computing, 2009, pp. 911-918.

[23] A. Wun et al. “A taxonomy for denial of service attacks in content-based publish/subscribe systems.” In ACM **DEBS,** 2007, pp.116-127.

[24] Y. Li et al. “Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach.” IEEE **TAC**, 60(10), pp.2831-2836.

[25] S. Misra et al. “Secure content delivery in information-centric networks: Design, implementation, and analyses.” In ACM **ICN** 2013, pp. 73-78.

[26] K. Zhou et al. “Security in cyber-physical systems: challenges and solutions.” **IJAACS** 2017, Vol. 68pp. 391-408.

[27] A. Humayed et al. "Cyber-physical systems security—A survey." IEEE Internet of Things Journal, 4(6).2017, pp. 1802-1831.

[28] P. Derler et al. “Modeling cyber-physical systems”. IEEE Proceeding, 100, 1 (2012), pp. 13–28.

[29] T. Bures et al. “Performance Modelling of Smart Cyber-Physical Systems,” in **ICPE** 2018, pp. 37-40.

[30] A. Easwaran et al. “A systematic security analysis of real-time cyber-physical systems.” **ASP-DAC**, 2017 pp. 206-213.

[31] K. A. Stouffer et al. “Guide to ICS security: SCADA, DCS, and another control system such as PLC,” NIST, Gaithersburg, MD, USA, Tech. Rep. Sp 800-82, 2011.

[32] C. Neuman, “Challenges in security for cyber-physical systems,” in Proc. DHS Workshop Future Directions Cyber-Phys. Syst. Security, Newark, 2009, pp. 22–24.

[33] H. Ge et al. “Analysis of Cyber-Physical Systems Security Issue via Uncertainty Approaches.” Advanced Computational Methods in Life System Modeling and Simulation. Springer, Singapore, 2017. Pp. 421-431.

[34] J. Wang et al. “Detecting time synchronization attacks in cyber-physical systems with machine learning techniques.” In IEEE **ICDCS** 2017, pp. 2246-2251.

[35] G. E. Suh et al. “Physically unclonable functions for device authentication and secret key generation,” in ACM **DAC**, 2007, pp. 9–14.

[36] M. Rahman et al, “An aging-resistant ro-puf for a reliable key generation,” IEEE Transactions on Emerging Topics in Computing, vol. PP, no. 99, 2015, pp. 335-348.

[37] O. Al Ibrahim et al. “Cyber-physical security using system-level pufs,” in **IWCMC**, 2011, pp. 1672–1676.

[38] L. Wei et al. “Boardpuf: Physical unclonable functions for printed circuit board authentication,” in IEEE **ICCAD**, 2015, pp. 152–158.

[39] M. Stipcevic et al.“True random number generators,” in Open Problems in Mathematics and Computational Science, pp. 275–315. Springer, 2014.

[40] M. T. Rahman et al. “Ti-trng: Technology independent true random number generator,” in ACM **DAC**, 2014, pp. 1–6.

[41] Y. Liu et al. “Impact assessment of net metering on smart home cyber attack detection,” in in ACM **DAC**, 2015, pp. 97:1- 97:6.

[42] A. Cardenas et al. “Challenges for securing cyber-physical systems.” Workshop on future directions in cyber-physical systems security. Vol. 5. 2009.

[43] F. Pasqualetti et al. “Design and operation of secure cyber-physical systems.” IEEE Embedded Systems Letters 7(1), 2015, pp. 3-6.

[44] D. I. Urbina et al. “Survey and new directions for physics-based attack detection in control systems”. US Department of Commerce, NIST, 2016.

[45] L.A. Tang et al. “Tru-alarm: Trustworthiness analysis of sensor networks in cyber-physical systems” In IEEE **ICDM**, 2010, pp. 1079-1084.

[46] K. Wan et al. “Context-aware security solutions for cyber-physical systems.” Mobile Networks and Applications 19(2), 2014, pp. 212-226.

[47] K. Wan et al. “Dependable context-sensitive services in cyber-physical systems.” In TrustCom, 2011, pp. 687-694.

[48] X. KoutsouKos, et al. "SURE: A Modeling and Simulation Integration Platform for Evaluation of SecUre and REsilient Cyber-Physical Systems." Proceedings of the IEEE 106(1), 2018, pp. 93-112.

[49] D. Gollmann et al. “Cyber-physical systems security: Experimental analysis of a vinyl acetate monomer plant.” Cyber-Physical System Security. ACM, 2015.

[50] F. Pasqualetti et al. “Design and operation of secure cyber-physical systems.” IEEE Embedded Systems Letters 7(1), 2015, pp. 3-6.

[51] J. Taylor et al. “Security challenges and methods for protecting critical infrastructure cyber-physical systems.” In IEEE **MoWNeT**, 2017, pp. 1-6.

[52] S. Peisert et al. “Designed-in security for cyber-physical systems.” IEEE Security & Privacy 12(5), 2014, 9-12.

[53] F. Pasqualetti et al. “Attack detection and identification in cyber-physical systems.” IEEE Transactions on Automatic Control 58(11), 2013, pp. 2715-2729.

[54] G. Sabaliauskaite et al. “Intelligent checkers to improve attack detection in cyber-physical systems.” In IEEE CyberC, 2013, pp. 27-30.

[55] A. L. Buczak et al, “A survey of data mining and machine learning methods for cybersecurity intrusion detection.” IEEE Communications Surveys & Tutorials 18(2), 2016, pp. 1153-1176.

[56] R. C. B. Hink et al. “Machine learning for power system disturbance and cyber-attack discrimination.” In IEEE ISRCS, 2014, pp. 1-8.

[57] S. Pan et al. “Developing a hybrid intrusion detection system using data mining for power systems.” IEEE Transactions on Smart Grid 6(6), 2015, pp. 3104-3113.

[58] Y. Zhang et al. “Health-CPS: Healthcare cyber-physical system assisted by cloud and big data." IEEE Systems Journal 11(1), 2017, pp. 88-95.

[59] A. Valdes at al. “Anomaly detection in electrical substation circuits via unsupervised machine learning.” In IEEE IRI, pp. 500-505.

[60] R. Mitchell et al. “Behavior rule specification-based intrusion detection for safety-critical medical cyber-physical systems.” IEEE Transactions on Dependable and Secure Computing 12(1), 2015, pp. 16-30.

[61] Sadeghi, Koosha, et al. “Toward parametric security analysis of machine learning based cyber forensic biometric systems,” in IEEE ICMLA, 2016, pp. 626-631.

[62] F. K. Lodhi et al. "Power profiling of microcontroller's instruction set for runtime hardware Trojans detection without golden circuit models." Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, 2017, pp. 294-297.

[63] F. K. Lodhi et al. “A self-learning framework to detect the intruded integrated circuits,” in IEEE ISCAS, 2016, pp. 1702-1705.

[64] A. Kulkarni, “Adaptive Real-time Trojan Detection Framework Through Machine Learning,” in **HOST**, 2016, pp. 120–123.

[65] A. Kulkarni et al. “SVM-based Real-time Hardware Trojan Detection for Many-Core Platform.” In Symposium on Quality Electronic Design, 2016, pp. 362–367.

[66] M. Tehranipoor and H. Salamani. trust-HUB, 2016. URL https://www.trust-hub.org/

[67] M. Melis et al. “Is deep learning safe for robot vision? adversarial examples against the icub humanoid.” arXiv preprint arXiv:1708.06939 (2017).

[68] N. Papernot et al. “The limitations of deep learning in adversarial settings.” European Symposium on Security & Privacy (EuroS&P), IEEE, 2016, pp. 372-387.

[69] X. Zhang et al. “Modular Learning Component Attacks: Today's Reality, Tomorrow's Challenge.” arXiv preprint arXiv:1708.07807 (2017).

[70] N Papernot et al. “Practical black-box attacks against machine learning.” Asia Conference on Computer and Communications Security. ACM, 2017, pp. 506-519.

[71] T. Gu et al.. “BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain.” arXiv preprint arXiv:1708.06733 (2017).

[72] M. Sharif et al. “Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition,” ACM CCS, 2016, pp. 1528-1540.

[73] R. Collobert et al. “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in ACM **ICML**, 2008, pp. 160-167.

[74] C. Szegedy et al. “Intriguing properties of neural networks.” arXiv preprint, arXiv:1312.6199 (2013).

[75] R. Gilad-Bachrach et al. "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in ACM **ICML**, 2016, pp. 201-210.

[76] H. Hosseini et al. “Attacking Automatic Video Analysis Algorithms: A Case Study of Google Cloud Video Intelligence API.” arXiv preprint, arXiv:1708.04301 (2017).

[77] Y. Liu et al. "Trojaning Attack on Neural Networks." (2017). Department of Computer Science Technical Reports. Paper 1781.

[78] Y. Liu et al. "Neural Trojans." arXiv preprint arXiv:1710.00942 (2017).

[79] K. Wang et al., “Jamming and Eavesdropping Defense in Green Cyber-Physical Transportation Systems using Stackelberg Game,” IEEE Transactions on Industrial Informatics, 2018.