Scientific Research Publishing

Journal of Information Security, 2025, 16(1), 149-157
https://www.scirp.org/journal/jis
ISSN Online: 2153-1242
ISSN Print: 2153-1234

Architecture to Secure Electrical Control System in Cyber-Physical System

Depeng Li

Department of Information and Computer Sciences, University of Hawaii at Manoa, Honolulu, HI, USA
Email: depengli@hawaii.edu

How to cite this paper: Li, D.P. (2025) Architecture to Secure Electrical Control System in Cyber-Physical System. *Journal of Information Security*, **16**, 149-157.
https://doi.org/10.4236/jis.2025.161008

Received: October 29, 2024
Accepted: December 31, 2024
Published: January 3, 2025

Abstract

It's possible for malicious operators to seize hold of electrical control systems, for instance, the engine control unit of driverless vehicles, from various vectors, e.g. autonomic control system, remote vehicle access, or human drivers. To mitigate potential risks, this paper provides the inauguration study by proposing a theoretical framework in the physical, human and cyber triad. Its goal is to, at each time point, detect adversary control behaviors and protect control systems against malicious operations via integrating a variety of methods. This paper only proposes a theoretical framework which tries to indicate possible threats. With the support of the framework, the security system can lightly reduce the risk. The development and implementation of the system are out of scope.

Keywords

Architecture, Control System, Framework

1. Introduction

In the past, cyber-attacks have not been introduced to our world. The malfunction of the control system has only been treated as a mechanical device failure. Customers' safety is rarely threatened by hackers via remote access cyber channels. Nowadays, electronic control systems are vulnerable to different kinds of attacks. Automobiles, for instance, can be manipulated by hackers through a variety of attack vectors [1]-[4]. In this paper, we would like to find the right operator when malicious ones are trying to take it over.

Previously, control systems such as automobiles, robots, and so on were handled only by onsite human operators. Gradually, control privileges have been granted to industrial autonomic control systems and, then, be granted to remote access cyber systems through network communication channels. This paradigm

shift offers incredible productivity and much more convenience. But a security concern has also been introduced: it is possible that autonomic control systems have been taken over by the wrong hand: the human operator can act maliciously, the autonomic control system can be infected, and the remote access cyber system can be compromised by hackers. The recent growth of misbehavior activities covers all possible aspects of Cyber Physical Systems (CPS) [1]-[4]. This paper's major motivation is that since adversaries could potentially launch attacks from every channel, we should mitigate the risks by identifying the malicious controllers and then preventing their attacks.

This paper intends to propose a theoretical framework that protects control systems against malicious operations launched by attackers, ranging from human operators to automation control systems to remote access cyber control systems. This paper, specifically, focuses on a system with more than one controller. Each of them can independently operate the control system. The main questions we want to address are 1) how to detect the malicious CPS controllers; 2) how to protect the control system after the malicious CPS controller is identified.

This paper cannot totally solve the problems above. But, to partially answer these questions, the new ideas are to 1) Propose the Physical, Human and Cyber Triad; and 2) Establish a quantitative framework that aims to develop a set of metrics which are used to assess and analyze the security condition of each controller instance. The driverless vehicle is used as a case study to verify this idea.

This paper can neither eliminate any attacks nor develop any real-time systems. Its goal is to lightly mitigate risks resulting from the aforementioned misbehaviors by proposing a framework.

2. Related Works

From the perspective of human operators, various guides, manuals or frameworks are published to regulate operations impacting safety, reliability and security [1] [5] [6]. As an example of an automation control system, the Supervisor Control and Data Acquisition (SCADA) control system was analyzed in areas of malware that could compromise critical infrastructure systems [7].

To study the problem of human-in-the-loop feedback control systems, modeling human behavior and incorporating the model into the formal feedback control system are briefly addressed [8]. A cyber-physical-social based security architecture (namely, IPM) studied three critical security perspectives: information, physical, and management [9]. Physical security is addressed by artificial immunity, and management security is achieved through social strategies. But, to the best of my knowledge, less attention is paid to studying critical topics: 1) effectively preventing malicious activities in CPS over the control system and 2) withdrawing the access right from misbehavior controllers have not been carefully studied.

To evaluate malicious activities and to prevent cyber-attacks, the reputation of hosts has been widely studied, which can detect, filter and block the misbehavior activities such as spam, unauthorized access control, etc. [10].

3. Theoretical Background

3.1. Group Key Scheme

In order to grant/withdraw the access right over the control system, we will leverage the Group Key Scheme [11], based on which the new CPS controller can join and the malicious CPS controller can be forced to leave. In this paper, each controller's activities will be viewed/audited by other peer controllers in real time.

The group key scheme we utilized includes two components: 1) efficient and reliable group key agreement; 2) virtual synchrony view. The former ensures that only the present group members contain the current group key. Malicious controllers will be expelled based on their misbehavior. The group key is a symmetric key shared among all current group members in such a way that the control commands or measured statuses will be encrypted by the group key during their transmission over control systems. A group key scheme should satisfy two privacies: 1) *Forward secrecy*: previous group members who know contiguous subsets of old group keys must not be able to discover subsequent group keys after they leave the group. 2) *Backward secrecy*: current group members who know a contiguous subset of current group keys must not be able to discover preceding group keys. The virtual synchrony view means that, if processes p and q install the same new view V in the same previous view V', then any message received by p in V' is also received by q in V'. So,

$$installs_{in(p,V,V')} \wedge install_{in(q,V,V')} \wedge receive_{in(p,m,V')} \rightarrow receives_{in(q,m,V')}. \quad (1)$$

Authentication and integrity will be provided for group keys to prevent attacks, e.g., Man-In-The-Middle attacks, etc. But we will not explain the details here.

3.2. Supervisory Control Theory

Our research tries to isolate the malicious CPS controllers from the control system by employing the supervisory control theory in which discrete state spaces and event-driven dynamics are widely used. From the viewpoint of discrete event systems, the control system under protection can be modelled as a plant to which the supervisory controllers send control actions. The control system is treated as the feedback control of dynamic systems, the specification of which is represented as finite-state automata over the set of discrete events. Any actions from the controllers will force the control system to make discrete changes and, consequently, generate a sequence of discrete events that can formulate transition-based, event-driven models. Research focusing on the models can analyze the security, safety, reliability, etc. [12]. Due to its capability to quickly detect and isolate malfunctions, the fault detection and diagnosis methodology synthesizing to discrete state event-driven dynamics is further studied to model the control system's activities and to further deduce controllers' misbehavior [13]. However, it is still a concern that they are sensitive to noises, which may easily lead to false alarms during normal operations.

3.3. Rejection for Cyber Misbehaviors

The reputation of a host [10] has been treated as a vital metric which measures the security condition of a host. Based on the reputation value, some systems construct a list of rejections. The purpose is to block/filter the inbound or outbound traffic sent from/forwarded to hosts in the list.

4. Overview of the Proposed System

4.1. Problem Descriptions

A control system, for example, a driverless vehicle, could be driven by a few controllers. At each time point t_i, the vehicle received a few control commands, one of which was issued by the controller in charge. Our decision is based on the current status of the vehicle (e.g. speed, engine brake, gas, light, etc.), context around the vehicle (e.g. other vehicles at the same lane or different lanes, road condition, etc.), and the mission of this trip. Then, we need to figure out how to evaluate which control command is safe. If the control command in charge is malicious, how can we find the right from others? To be more generic, let us assume the controller may be physical, cyber or human operators, namely, $O=\{O_x\}$, where $x=1,\cdots n$. Let P denote the control system, say, driverless vehicle which is fed with a number of control commands from different controllers at every time point. Therefore, at a sequence of time points, $\{t_1,t_2,\cdots,t_m,\cdots\}$, there are a series of control commands $\left\{\{c_{O_1},c_{O_2},\cdots\}_{t_1},\{c_{O_1},c_{O_2},\cdots\}_{t_2},\cdots\{c_{O_1},c_{O_2},\cdots\}t_m\cdots\right\}$. At each time point, the mobile control system could be at a status, S_k where $S_k\in S$ and $S_k=[M]_{i\times j}$ and i,j,m are integers. Note that $[M]_{i\times j}$ represents a block of key parameters which are measured from object, P. The *problem* we would like to address is that, at time point t_i, we need to select a control command $c_x\in\{c_{O_1},c_{O_2},\cdots\}_{t_i}$ in a sense that c_x will lead driverless vehicle from status S_{t_i} to $S_{t_{i+1}}$ in which the condition $S_{t_{i+1}}$ is safe. The challenge is (1) how to validate c_x in case of the context set $\varphi(t_i)=\{\varphi_1,\varphi_2,\cdots\varphi_n\cdots\}$ and the mission M where φ_i is one element of context, (2) how to drop the malicious commands and (3) how to punish the corresponding unsafe controller.

4.2. Physical-Human-Cyber Triad

As depicted in **Figure 1**, our paper proposes a new physical, human and cyber triad system, which is comprised of three components (sketched as circles), each denoting one type of controller.

The system under protection is represented by the "control system" blocks (sketched as rectangles), which abstract a general model with such instances as modern vehicles, unmanned robots, etc. The solid, directional links connecting controller elements and the control system blocks denote the capability of the control privilege, and the dashed links represent functions of collecting statuses from control systems.

A controller operates control systems through three channels: the human operators drive the handle, e.g., function lever, the automation control systems

execute embedded operational instructions in firmware, and the remote access cyber systems forward control commands that are encapsulated in packets through cyber communication channels.

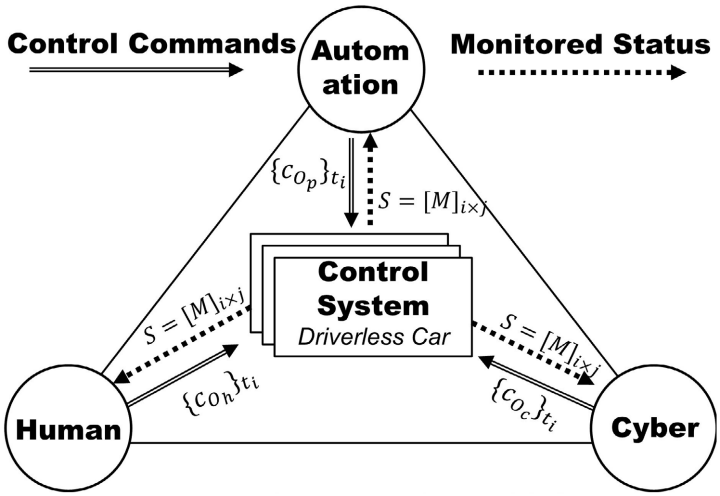

Figure 1. Physical, Human and Cyber (PHC) Triad.

This paper assumes that any controller in a physical, human and cyber triad could act maliciously since, even if not hostile at the beginning, it may potentially be compromised by attackers in future. The paper will use this conceptual architecture to offer a quantitative, abstract model to assess the compromised controller based on their misbehavior activities and to further provide a mechanism which can withdraw the control privileges after the detection of corresponding attacks.

4.3. Architecture

This paper seeks to mitigate the threat introduced by the following problems for each physical, human and cyber controller: 1) Define a variety of metrics to model the misbehaviors over the control systems; 2) Construct the quantitative framework to evaluate the security conditions of the control system based on the captured operational commands and the corresponding results; and 3) Define reputation index thresholds in the quantitative framework and withdraw adversary controller's control privilege when its reputation index is beyond the threshold.

Note that this paper only proposes the framework rather than studying the details due to the limited space.

So, the following work will not be considered in this paper. But they can be accomplished in our future work:

1) Fulfill numerous data collection, data separation, and data analyses over not only control commands instructed by all physical, human and cyber controllers but also the status result data measured over the control system for both before and after the control commands are executed;

2) Construct a generic model to measure a set of essential metrics which can

not only evaluate the security condition of control systems but also assess other critical parameters such as reliability, safety and so on;

3) Design the quantitative framework indicating the security level for each controller, namely, the autonomic controller, human operators, and remote access cyber controllers. Security conditions will be assessed by discrete-state event-driven fault-diagnosis theories [14] [15], by human operation manuals [1] [5] [6], by finite-state machines, and by reputation systems [10], and;

4) Withdraw the malicious controllers' control privileges via group key schemes [11] after attacks are detected.

5. Our System

We outline our system here, but the detailed implementation will be our future task. As depicted in Figure 2, our architecture includes four layers.

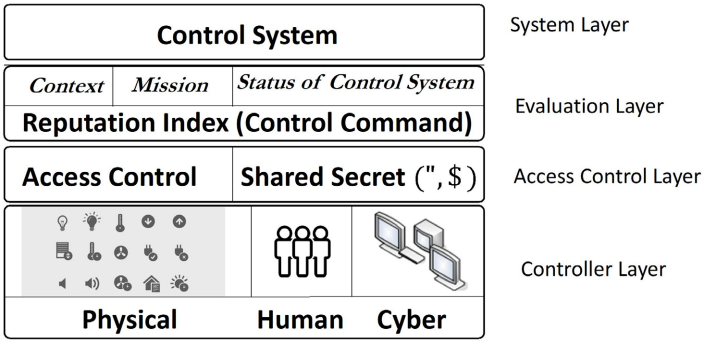

Figure 2. Architecture of our system.

In the control layer, control commands are issued by physical (autonomic control system), remote access cyber, and human operators.

In the quantitative layer, a set of metrics is established for each controller in the physical, human and cyber triad. Each controller's metric indicator decides whether this controller instance is malicious or not. In the access control layer, the detected malicious controller instance will be expelled from the group key agreement in such a way that a new group key is generated, and the expelled group member cannot access the new group key. So, there is no chance for it to send its control commands to the control system.

In the system layer, the control system decrypts the ciphertext command by using the new group key and then executes the command. Malicious/illegal controller instances' ciphertext cannot be correctly decrypted since they cannot hold the current group key. Just notice that the controller instance can access the control systems by utilizing wearable devices such as a master card/token.

Note that after executions, the monitored statuses of the control system will be used to analyze the behavior of each control command. If the result is unsafe/negative or even an alarm is triggered, the negative value will be reflected in the metric indicators of corresponding controllers. The loopback of the control system is critical for this paper as it demonstrates each controller's behavior, which impacts

its reputation index.

5.1. Subsystem

Capture Control Commands and Statuses:

We collect the control commands that are issued by physical, human, and cyber controllers, as well as a finite sequence of statuses of control systems. Those raw data will be de-noised, stored, classified, filtered and evaluated through a set of pre-processing operations or algorithms.

Autonomic control system (*P*): we first map a sequence of observable events $E=\{e_1,e_2\cdots,e_n\}$ to a set of control actions $C=\{c_1,c_2\cdots,c_n\}$. Function $F:E\times C=\{e_i\}\times\{c_j\}$. Our intrusion detection system is defined as I (a fault-diagnosis function). The control loop $A\to F$ (where A: finite-state automation) is named as a potential attack if the action c_j is misbehavior and if the e_i is one of the fault states of the control system. Combining the active fault-diagnosis theory and the finite-state automation method, the intrusion detection I could abstract the control system as discrete-state event-driven dynamic and identify the misbehavior or attacks as an active fault event. If a fault even occurs, the corresponding action c_j will be traced back to its controller, whose reputation index, in turn, will be impacted.

Human Operator (*H*): wrong human operations can negatively impact the control systems, which are shown as different kinds of symptoms. We can collect a number of symptoms led by the misbehavior which will be treated as complementary to some well-known guides or user manuals [1] [5] [6]. They altogether can correct the mis-operation and stop adversaries.

Cyber remote control (*C*): we will analyze the malicious activities launched from remote access cyber controller through the method for data collection and measurement. Its result reflects the reputation index of each specific remote host, based on which the list of rejection can be constructed.

Misbehavior Abstraction, Profiling and Modeling:

While analyzing control commands, we aim to identify, profile, model and filter attacks based on a formal method via utilizing finite-state machine, namely Φ, which is listed below:

$$S\times C\xrightarrow{\Phi}S\times O \tag{2}$$

$$\Phi(S,C)=\Phi\left(S,\{c_1,c_2,\cdots c_n\}\right) \tag{3}$$

$$=\Phi\left(\Phi(S,c_1),\{c_2,\cdots c_n\}\right) \tag{4}$$

where

S is the state,

C is the control command sequence,

O is the output,

c_n is each control command.

Record Misbehavior:

Regarding each control command c_i on the list of rejection, the controller which

sends out c_b will be impacted for its reputation index.

Access Control (group key scheme):

The group key will be *rekeyed*, which can guarantee both forward privacy and backward privacy. Thus, the expelled controller cannot get the subsequent group keys. The control system is assumed to always hold the current group key and therefore can decrypt the ciphertext. The other advantage is that peer and legal controllers can audit others' activities in order to prevent the misbehavior via group key.

5.2. Case Study

Let us take a driverless car as an example: at a time point t_i, a few controllers, O_h, O_p, and O_c represent the human driver, the physical automation control device in the vehicle and the remote access cyber control program, respectively. They issue control command $\left\{c_{O_h}, c_{O_p}, c_{O_c}\right\}_{t_i}$. In our paper, since each of O_h, O_p, and O_c could be malicious, we will choose the right control command O_r for driverless vehicle P based on the current context set $\varphi(t_i)$ and the vehicle statuses, $S_k = [M]_{i\times j}$. The context set $\varphi(t_i)$ could include the road traffic, the speed of other vehicles in front/behind the vehicle or at neighbor lanes, the pedestrian on the road, the condition of the road, parking space, gas level, map service, and so on. In a word, the context, together with the current status of the vehicle as well as the mission of the trip, can decide whether the vehicle is safe or not. Our future work includes both simulation and real-world experiments.

6. Conclusion

In this paper, the proposed solution can decide which control command is malicious and how to select the right one from the others. We also record the malicious commands/misbehavior, which will impact the controller's reputation index. The scheme could lightly mitigate the attacks from a variety of control vectors for CPS. Furthermore, this paper not only provides a framework to identify the misbehavior of the adversary controllers in the Physical, Human and Cyber triad, but also briefly explains how to construct a dataset. The dataset contains the misbehavior based on the specific status of the driverless vehicle and the corresponding context set. The goal of this paper is the development of a framework for protecting control systems against malicious operations. We cannot completely eliminate the attack but rather mitigate the risk resulting from misbehaviors.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] How to Prevent a Robot Rebellion. http://www.createthefuturecontest.com/

[2] India Blackouts (2012) Report on Grid Disturbance on 30th & 31st July 2012. http://www.cercind.gov.in/2012

[3] (2014) US Military Begins Research into Moral, Ethical Robots, to Stave off Skynet-Like Apocalypse. http://www.extremetech.com/

[4] Checkoway, S., *et al*. (2011) Comprehensive Experimental Analyses of Automotive Attack Surfaces. *USENIX Security Symposium* 2011, San Francisco, 8-12 August 2011, 1-16.

[5] Hawaii Driver's Manual. https://m.driving-tests.org/hawaii/hi-dmv-drivers-handbook-manual/

[6] Substation Operation and Maintenance. https://www.energy-consult.net/en/services/operation-maintenance-of-substations/

[7] Cardenas, A.A., Amin, S. and Sastry, S. (2008) Secure Control: Towards Survivable Cyber-Physical Systems. 2008 *The* 28*th International Conference on Distributed Computing Systems Workshops*, Beijing, 17-20 June 2008, 495-500. https://doi.org/10.1109/icdcs.workshops.2008.40

[8] Munir, S., Stankovic, J.A., Mike Liang, C.-J. and Lin, S. (2014) Cyber Physical System Challenges for Human-in-the-Loop Control. *The* 8*th Workshop on Feedback Computing*, *USENIX Security* 2014, San Diego, 20-22 August 2014, 1-4.

[9] Ning, H. and Liu, H. (2012) Cyber-Physical-Social Based Security Architecture for Future Internet of Things. *Advances in Internet of Things*, **2**, 1-7. https://doi.org/10.4236/ait.2012.21001

[10] Pathak, A., Qian, F., Hu, Y.C., Mao, Z.M. and Ranjan, S. (2009) Botnet Spam Campaigns Can Be Long Lasting. *ACM Sigmetrics Performance Evaluation Review*, **37**, 13-24. https://doi.org/10.1145/2492101.1555352

[11] Li, D. and Sampalli, S. (2008) A Hybrid Group Key Management Protocol for Reliable and Authenticated Rekeying. *International Journal of Network Security*, **6**, 228-270.

[12] Girard, A. and Pappas, G.J. (2007) Approximation Metrics for Discrete and Continuous Systems. *IEEE Transactions on Automatic Control*, **52**, 782-798. https://doi.org/10.1109/tac.2007.895849

[13] Luzar, M., Czajkowski, A., Witczak, M. and Korbicz, J. (2012) Actuators and Sensors Fault Diagnosis with Dynamic, State-Space Neural Networks. 2012 17*th International Conference on Methods & Models in Automation & Robotics* (*MMAR*), Miedzyzdroje, 27-30 August 2012, 196-201. https://doi.org/10.1109/mmar.2012.6347889

[14] Banerjee, A., Venkatasubramanian, K.K., Mukherjee, T. and Gupta, S.K.S. (2012) Ensuring Safety, Security, and Sustainability of Mission-Critical Cyber–Physical Systems. *Proceedings of the IEEE*, **100**, 283-299. https://doi.org/10.1109/jproc.2011.2165689

[15] Li, D., Aung, Z., Williams, J.R. and Sanchez, A. (2014) Efficient and Fault-Diagnosable Authentication Architecture for AMI in Smart Grid. *Security and Communication Networks*, **8**, 598-616. https://doi.org/10.1002/sec.1006