

FAILURE-ON-DEMAND PROBABILITY AND MALFUNCTION RATE ESTIMATION IN NUCLEAR POWER PLANT CYBER-PHYSICAL SYSTEMS

Wei Wang¹, Francesco Di Maio¹, Enrico Zio^{1,2}

¹Energy Department, Politecnico di Milano, Via La Masa 34, 20156 Milano, Italy

²Chair on System Science and the Energy Challenge, Fondation Electricite' de France (EDF), CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France
wei.wang@polimi.it, francesco.dimaio@polimi.it, enrico.zio@polimi.it

Nuclear Power Plants (NPPs) are making increasing use of digital Instrumentation and Control (I&C) systems, which makes them Cyber-Physical Systems (CPSs). In CPSs, cyber and physical processes are dependent and interact with each other: sensors, actuators, communication and computational units are all interconnected to realize real-time monitoring, dynamic control and decision support, for normal operation as well as in case of accidents. However, an emerging concern is that the use of computer-based technologies might increase the exposure to failures and accidents, providing new channels for their initiation and propagation. System integrity can be, indeed, affected by hardware component failures, human errors, communication malfunctions and software errors, but also compromised by security breaches and cyber attacks. These latter can, in practice, be misclassified as components failures-on-demand and malfunctions, hiding malicious cyber attack threats.

In this study, we investigate and analyze the modelling of stochastic failures in CPSs, with the purpose of estimating reference values of failure-on-demand probabilities and malfunction rates for their components, so that any difference with statistical estimates from field data collected on the real CPS values can be used to detect malicious attempts at altering the safety of a NPP. A digital I&C system of a NPP is taken as illustrative case study, in which stochastic components failures resulting in different system responses are analyzed and Fault Tree (FT) analysis and Markov Chain Model (MCM) are taken as approaches to estimate the reference failure-on-demand probabilities and malfunction rates.

I. INTRODUCTION

A Cyber-Physical System (CPS) features a tight combination of (and coordination between) the system computational units and physical elements. The integration of computational resources into physical processes is aimed at adding new capabilities to stand-

alone physical system and realize real-time monitoring, dynamic control and decision support during normal operation as well as in case of accidents. In CPSs, cyber and physical processes are dependent and interact with each other through feedback control loops (e.g., embedded cyber controllers monitor and control the system physical variables, whilst physical processes affect, at the same time, the monitoring system and the computation units by wired or wireless networks (Refs. 1, and 2)). The benefit such self-adaptive capability makes CPSs to be increasingly operated in transportation, energy, medical and health-care, and other applications (Refs. 2, 3, and 4). In the context of nuclear energy, digital Instrumentation and Control (I&C) systems of Nuclear Power Plants (NPPs) can be considered CPSs being, nowadays, strongly relying on computer-based functions for enhancing NPP availability.

Cyber controllers have been shown to benefit from the adequate use of information related to (1) environmental conditions (which plays an important role in affecting the system dynamics and should be measured and adaptively integrated into the cyber real-time monitoring and control in an intelligent manner (Ref. 5)), (2) periodically updated database of parameters (for keeping up-to-date the CPS settings (Ref. 6)), (3) new interaction modalities between human and system user interfaces (leading to more flexible system operability from the human perspective (Ref. 7)), and (4) computer-based networks status (to enhance the network connectivity and remote control, communicate with sensing data, and coordinate over constraint environment (Ref. 8)).

It is common practice that Probabilistic Safety Assessment (PSA) analysts generally focus on hardware component failures, human errors, communication malfunctions and software errors to assess the consequences of accidental hazards. However, the increase of computer-based technologies calls not only to account in PSA for accidental hazards that may affect physical systems, but also for threats to system integrity, such as interrupted communication between the cyber

system and the external environment due to malicious attacks through networks. Such cyber threats, initiated in the cyber domain but only being manifested in the physical domain, can, in practice, be misclassified as component stochastic failures by risk analysts, hiding their malicious effects on the system security (Refs. 9, and 10). In this sense, malicious attempts aimed at altering the CPS normal operation should be detected and distinguished from component stochastic failures. This can be done either by scenario processing (i.e., modeling the malicious cyber events and their manifestation on the physical domain, affecting, in turn, both cyber and physical properties of the CPS), or by estimation of reference values of probabilistic metrics (e.g., failure-on-demand probabilities and malfunction rates) of the CPS components, so that any difference with statistical estimates from field data collected on the real CPS values can be used to detect malicious attempts at altering the safety of a NPP.

In this study, as well as in Refs. 11, and 12 where other techniques are explored, we follow the latter alternative and investigate, analyze and model CPSs failures, with the purpose of estimating reference values for the failure-on-demand probabilities and the malfunction rates. A digital I&C system of a NPP is taken as illustrative CPS case study, in which components stochastic failures resulting in different system responses are analyzed and Fault Tree (FT) analysis and Markov Chain Model (MCM) are taken as approaches to estimate the reference failure-on-demand probabilities and malfunction rates.

The remainder of the paper is organized as follows. Section II clarifies the need of including security (i.e., cyber attacks) into risk assessment, traditionally considering only safety aspects (i.e., components stochastic failures). In Section III, failures-on-demand and malfunctions of a typical CPS of a NPP are described. Reference failures-on-demand probabilities and malfunctions rates are estimated for the components of the digital I&C CPS considered as case study in Section IV. Section V draws the conclusions of the work and provides suggestions on the benefits of such integrated safety and security analysis.

II. RISK ANALYSIS OF CPSs: THE INTEGRATION OF SAFETY AND SECURITY ANALYSIS

Risk is the likelihood of a hazard (or threat) to escalate from potential to real damage. This definition envelopes both safety and security aspects (Ref. 13). In safety analysis, the hazards generated from the components stochastic failures can result in unacceptable consequences on the system physical processes, whereas, security analysis focuses on malicious threats (in both physical and cyber domains).

Risk analysis has, therefore, to closely address both safety and security aspects, being these sharing many interdependencies and effects on the system. The challenge is that, even if components stochastic failures and cyber threats are distinct and diverse failure mechanisms, they can lead to same consequences on the system physical processes, and, therefore, cannot be distinguished during the system operation by the risk analysts (that means, on the other hand, that a malicious activity might be undergoing without being effectively detected). The integrity of safety and security CPS risk analysis is expected to allow identifying both accidental hazards rooting in component/system and cyber-related threats coming from external stimuli, to distinguish them from each other and to design effective countermeasures and protections.

II.A. Accidental Hazards

In CPSs, integrity and functionality can be compromised by two alternative classes of failures (i.e., hardware and software) where both might be initiated by human errors (unintentional).

(1) Hardware failures

During the CPS operation, its embedded hardware components (i.e., sensors and actuators) can fail due to process and operational conditions that affect the way components interact with each other, aging that affects the process dynamics of the hardware failure behaviors, and degradation that generates multiple failure modes which affect the system response to different stimuli (Ref. 14). Hardware failures can lead to two types of misoperations of the CPSs:

- Failure-on-demand, that consists in failing to trigger protections or execute proper control strategies (when demanded), leading the system to reach (unknown) accidental scenarios;
- Malfunction, that consists in spuriously triggering protections (e.g., unintentional shutdown) or executing mistaken control strategies.

Failure-on-demand probability and malfunction rate should be estimated to support the decision-maker on the system periodic test and maintenance strategy to be enforced for minimizing the occurrence of the hardware stochastic failures during the system operational life.

(2) Software errors

In CPS, computational units connect actuators with sensors and can take self-adaptive and coordinative control decisions, based on the predefined logics and algorithms (Ref. 15). The therein embedded software, for example, Proportional-Integral-Derivative (PID) used as feedback controller in CPSs can retroact to actuators the actions to be undertaken for responding to the changes of physical parameters. Software errors (generated from the inadequate specification, incomplete testing scope and algorithm/logic failures) are latent and hidden in the

software design and triggered only when context modifications are to be met. For example, the PID inadaptability to vary its parameters to variable physical parameters can affect the robustness of CPS control rules, but cannot be disclosed unless the physical parameters are met during the PID operation.

In further detail, also software failures can be classified as passive failures (failure-on-demand) and active failures with spurious actuations (malfunction) (Ref. 16). Failure-on-demand probabilities and malfunction rates of software have gained increasing attentions in risk community in latest years (Refs. 11, and 17).

II.B. Cyber Threats

The CPSs connections with internet might increase the exposure of CPSs to accidents, such as cyber attacks to cyber controllers, databases, networks and human-system interfaces that can result in the interruption of system integrity. Malicious activities can be categorized into Denial of Service (DoS) attacks, false data injection attacks (e.g., packet/data modification), network scan & sniffing attacks, integrity attack (e.g., through malware contagion) and, illegal command execution (Refs. 10, 12, and 18). They are usually initiated in the cyber domain through local or remote accesses, mimicking the components stochastic failures but isolating the connectivity between cyber and physical systems, to leave the physical process uncontrolled toward severe consequences. Indeed, despite distinct properties and different occurrence frequencies, a component stochastic failure and a cyber attack can lead to same consequences on the system integrity. For example, under a certain emergent condition when the system shutdown is demanded, both the actuator failing to trigger and the attacker intercepting the shutdown command to replace with normal information result in the same system accidental consequences, such that risk analysts can neither well detect the causes of the failures nor take right decisions to respond to the system accidents. In this sense, malicious attempts at altering the normal operation and in case of emergency, if neglected, can be misclassified as component misoperations, i.e., failure-on-demand and malfunction.

In practice, the fact that mechanisms and properties of the cyber threats are different from those of the accidental hazards might result to be of help for distinguishing them. Components stochastic failures can be detected by comparing the field data collected from the deployed redundancies of sensors or functional relationships among correlated quantities (Ref. 19); whereas, malicious cyber attacks are usually intermittent and some originally correlated physical variables might be found to be surprisingly non-correlated owing the attack (Ref. 10).

In what follows, instead of resorting to such scenario processing approach, we propose to estimate reference values (prior beliefs) for the failure-on-demand probabilities and the malfunction rates of a CPS, only based on hardware stochastic failures, to be used to detect any difference with statistical estimates from field data continuously collected on the real CPS and, finally, to distinguish malicious cyber attacks from components stochastic failures.

III. CASE STUDY

As shown in Fig. 1, the digital I&C system of a NPP as illustrative CPS case study is composed of two independent sensors (i.e., S-A and S-B), two channels of the cyber controller and one actuator (i.e., Reactor Trip Breaker (RTB)). Each channel of the cyber controller consists of the computational units of one independent Bistable Processor Logic (BPL, i.e., BPL-A and BPL-B) and one Local Coincidence Logic (LCL, i.e., LCL-A and LCL-B). If any of the two redundant measured signals exceeds a safety threshold value $V_{threshold}$, a Partial Tripping Signal (PTS) is measured from the corresponding BPL, e.g. a PTS from BPL-A is measured because S-A exceeds the tripping value. The signal processing proceeds only if both channels produce the PTS: each PTS from a BPL is sent to both LCL-A and LCL-B, which process information by an “AND” gate. In other words, an Emergency Shutdown Signal (ESS) is produced only when receiving two PTSs from different BPLs; ESSs, then, activates the RTB, when at least one ESS is triggered, i.e., the information is processed by an “OR” gate. Once the RTB is activated, the supported systems connected with RTB in the physical domain (e.g., power supply system, control rod drive mechanism, etc.) come into use to shutdown the reactor (Ref. 14).

The system functionality can be affected by sensors stochastic performance degradations and failures, RTB failures and the intermediate processor design errors. Besides these failures resulting in the system failures-on-demand and malfunctions, potential cyber attacks can threaten the system functionality with the same consequences of stochastic failures. Without loss of generality, two potential cyber threats scenarios are mentioned:

- False data injection: measured signals of both S-A and S-B exceeding $V_{threshold}$ requiring the RTB trigger are replaced with a string of normal information manipulated through contagion of malware, resulting in (intentional and malicious) RTB failure-on-demand. PTS cannot be measured from BPL-A and, as a result, the RTB will not be activated.
- Illegal command execution: a spurious command is injected into an intermediate processor (e.g., the LCL-A spuriously generates an ESS) resulting in

an altered control logic and a (intentional and malicious) RTB malfunction.

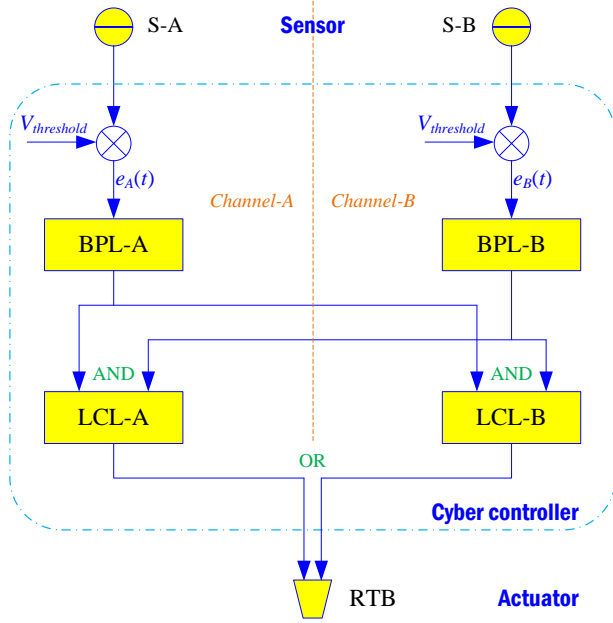


Fig. 1. The digital I&C system.

IV. ESTIMATION OF REFERENCE FAILURE-ON-DEMAND PROBABILITY AND MALFUNCTION RATE

Reference values of the failure-on-demand probability and the malfunction rate of the digital I&C system, are estimated based on the modeling of the hardware stochastic failures, to be used, in security analysis, to detect any difference with statistical estimates from field data collected on the system operation. The control scheme of Fig. 1 is designed without any time-dependent logic and dynamic feedback, so conventional modeling approaches are, herein, used for the illustrative identification of component stochastic failures, despite more comprehensive insights of accidental evolution obtained from dynamic modeling approaches.

IV.A. Estimation of the Failure-on-demand Probability

FT analysis is taken as the approach to estimate the reference failure-on-demand probability, in which the event “No shutdown signal is sent from the digital I&C system when demanded” as the top event. The fault tree is shown in Fig. 2, where common-cause failure between the LCLs is considered (β is equal to 0.1). Failure data of the basic events are taken from a public database (Ref. 20) and reported in TABLE I.

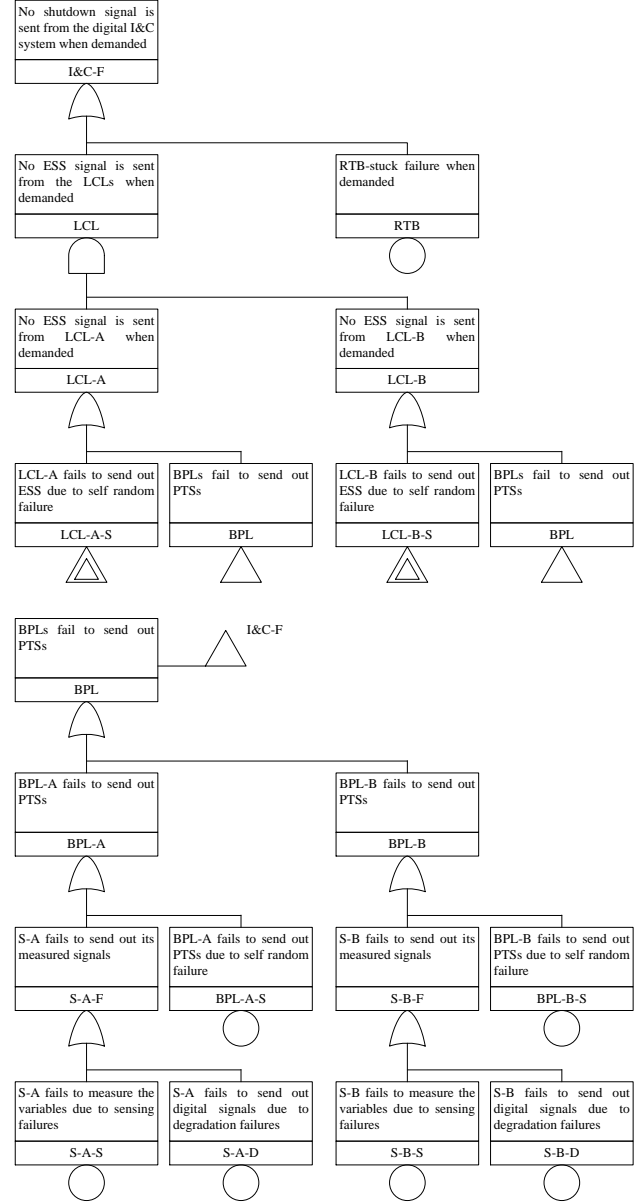


Fig. 2. FT for the system failure-on-demand probability estimation.

TABLE I. Components Failure Data (Ref. 20)

ID	Failure probability (/d)
S-A-S	5.85E-05
S-B-S	
S-A-D	5.85E-05
S-B-D	
BPL-A-S	5.44E-04
BPL-B-S	
LCL-A-S	1.60E-04
LCL-B-S	
RTB	1.54E-05

Based on Boolean logic quantification, the reference failure-on-demand probability (i.e., top event probability) results to be equal to $1.562E-6/d$.

IV.B. Estimation of the Malfunction Rate

A hybrid FT and MCM approach is used to estimate the malfunction rate. Minimal Cut Sets (MCSs) are obtained from the system malfunction FT, to present the system intrinsic properties and malfunction logics. Then, each of the obtained MCSs is modeled by a MCM, to estimate the MCSs time-dependent absorption state probabilities.

As shown in Fig. 3, the event “Shutdown signal is spuriously sent out from the digital I&C system” is identified as the top event, and the malfunction FT is built from top event to basic events according to the system configuration.

Table II. Estimation of the Malfunction Rates of the MCSs

Type	MCS	Malfunction rate/hr
1	{LCL-A-S}	4.450E-06
	{LCL-B-S}	4.450E-06
	{RTB}	1.169E-07
2	LCLs-common cause	4.994E-07
3	{S-A-S, S-B-S}	8.388E-10
	{S-A-D, S-B-D}	8.388E-10
	{BPL-A-S, BPL-B-S}	4.951E-09
4	{S-A-S, S-B-D}	8.388E-10
	{S-A-D, S-B-S}	8.388E-10
	{S-A-S, BPL-B-S}	2.038E-09
	{BPL-A-S, S-B-S}	2.038E-09
	{S-A-D, BPL-B-S}	2.038E-09
	{BPL-A-S, S-B-D}	2.038E-09

TABLE II lists the MCSs of the FT of Fig. 3. The malfunction rate, ω , for each MCS can be estimated by resorting to the system failure intensity quantification (Ref. 21):

$$\omega = \sum_{i \in S} p_i(t) \cdot \lambda_{i \rightarrow F} = \frac{1}{T_I} \int_0^{T_I} p_i(t) \cdot \lambda_{i \rightarrow F} dt \quad (1)$$

where, S is the set of success states of the MCM; F is the set of malfunction states; $p_i(t)$ is the probability of the system being in the success state i at time t ; $\lambda_{i \rightarrow F}$ is the transition rate of leaving success state i towards any failure state. TABLE III lists the transition rates $\lambda_{i \rightarrow F}$ taken from a public database (Refs. 20, and 22).

Realistically, for any component, a periodic testing interval is chosen equal to $T_I = 1$ year and mean

maintenance time to $T_R = 8h$, resulting in a repair rate μ for all equal to:

$$\mu = \frac{1}{\frac{T_I}{2} + T_R} = 2.28E-4/hr \quad (2)$$

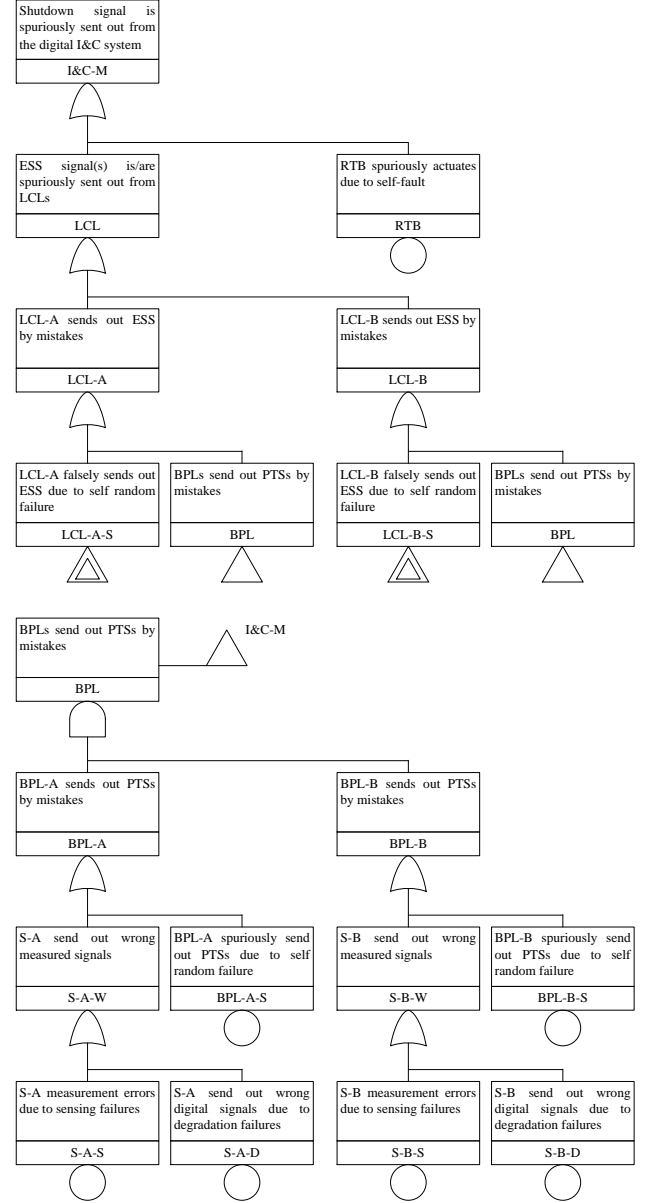


Fig. 3. FT for the system malfunction rate estimation.

We take the MCM of the MCS {S-A-S, BPL-B-S} as an illustrative example, to estimate the MCS malfunction rate. Fig. 4 shows the MCM of the MCS {S-A-S, BPL-B-S}, which consists of two different basic events. Four nodes are identified in the MCM, i.e., the functioning state “0”, the failure state “F”, and the intermediate states

“1_{SS}” and “1_{BS}” denoting S-A and BPL-A fails, respectively.

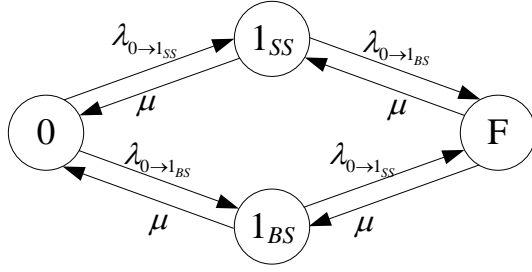


Fig. 4. Malfunction Markov chain model of the MCS {S-A-S, BPL-B-S}.

TABLE III. Transition Rates (Refs. 20, and 22)

Symbol	ID	Failure Mode	Failure Rate (/hr)
$\lambda_{0 \rightarrow 1_{SS}}$	S-A-S	S-A measurement error failure rate	4.11E-07
	S-B-S	S-B measurement error failure rate	
$\lambda_{0 \rightarrow 1_{SD}}$	S-A-D	S-A wrong digital output failure rate	4.11E-07
	S-B-D	S-B wrong digital output failure rate	
$\lambda_{0 \rightarrow 1_{BS}}$	BPL-A-S	BPL-A malfunction failure rate	1.00E-06
	BPL-B-S	BPL-B malfunction failure rate	
$\lambda_{0 \rightarrow 1_{LS}}$	LCL-A-S	LCL-A malfunction failure rate	5.00E-06
	LCL-B-S	LCL-B malfunction failure rate	
$\beta_{0 \rightarrow L_{CLS}}$	LCL-A-D & -B-D	LCLs common cause factor	1.00E-01
$\lambda_{0 \rightarrow 1_R}$	RTB	RTB malfunction failure rate	1.17E-07
μ	-	Repair rate	2.28E-04

Then, the malfunction rate during one periodic testing interval is,

$$\omega_{\{S-A-S, BPL-B-S\}} = \frac{1}{T_I} \int_0^{T_I} \lambda_{BS} \cdot p_{1 \rightarrow 3}(t) + \lambda_{SS} \cdot p_{2 \rightarrow 3}(t) dt = 2.038E-09 / hr$$

and results for all the MCSs are as shown in TABLE II (last column).

The calculated reference malfunction rate (that accounts for all MCSs of the digital I&C system) turns to be equal to 9.533E-6/hr. This enables analysts to compare the field data with the calculated reference malfunction rate: assume that numerous RTB failures-on-demand are recorded in a short period of time (e.g., one month), that

makes the failure-on-demand probability largely exceeding the reference value of 1.562E-6/d, and/or unexcepted shutdown occurrence making the malfunction rate largely exceeding the reference value of 9.533E-6/hr, these evidences should raise the analyst attention with respect to false data injected into S-A and S-B sensors databases and external illegal command attacks, respectively.

V. CONCLUSIONS

In this study, the twofold characteristics of CPS risk assessment has been pointed out. In particular, the necessity of investigating CPS risk with respect to both components stochastic failures and cyber threats has been highlighted. Concerns have been raised regarding the possibility of misclassification of cyber attacks as component failures-on-demand and malfunctions. To avoid the misclassification, malicious attempts aimed at altering the CPS normal operation should be detected and distinguished from components stochastic failures.

A probabilistic approach is here undertaken to address the problem. We propose to estimate reference values (prior beliefs) for the failure-on-demand probabilities and the malfunction rates of a CPS, only based on hardware stochastic failures, to be used to detect any difference with statistical estimates from field data continuously collected on the real CPS and, finally, to distinguish malicious cyber attacks from components stochastic failures.

The approach has been tested on a digital I&C system of a NPP. FT analysis and MCM are taken as approaches to estimate the reference failure-on-demand probability and malfunction rate of the system. Quantitative estimates allow for the detection of anomalous frequencies of system misoperations that would be initiated from the malicious attempts at altering the safety of the NPP.

REFERENCES

1. K. D. KIM and P. R. KUMAR, Cyber-Physical Systems: A Perspective at the Centennial. *Proceedings of the IEEE*, 100(Special Centennial Issue), 1287-1308 (2012).
2. E. A. LEE, Cyber Physical Systems: Design Challenges. In *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, p. 363-369, IEEE (2008).
3. S. K. KHAITAN and J. D. MCCALLEY, Design Techniques and Applications of Cyberphysical Systems: A Survey. *IEEE Systems Journal*, 9(2), 350-365 (2015).
4. J. M. BRADLEY and E. M. ATKINS, Optimization and Control of Cyber-Physical Vehicle Systems. *Sensors*, 15(9), 23020-23049 (2015).

5. W. WANG, F. DI MAIO and E. ZIO, Hybrid Fuzzy-PID Control of a Cyber-Physical System Working Under Varying Environmental Conditions. *Submitted to IEEE Transactions on Industrial Electronics*. (2016).
6. K. LIU et al., Temporal Data Dissemination in Vehicular Cyber-Physical Systems. *IEEE Transactions on Intelligent Transportation Systems*, 15(6), 2419-2431 (2014).
7. V. PAELKE and C. RÖCKER, User Interfaces for Cyber-Physical Systems: Challenges and Possible Approaches. *In International Conference of Design, User Experience, and Usability*, p. 75-85, Springer International Publishing (2015, August).
8. S. ALI et al., Network Challenges for Cyber Physical Systems with Tiny Wireless Devices: A Case Study on Reliable Pipeline Condition Monitoring. *Sensors*, 15(4), 7172-7205 (2015).
9. J. ZALEWSKI et al., A Framework for Measuring Security as a System Property in Cyberphysical Systems. *Information*, 7(2), 33 (2016).
10. M. S. RAHMAN et al., Multi-Agent Approach for Enhancing Security of Protection Schemes in Cyber-Physical Energy Systems. *IEEE transactions on industrial informatics*, 1-10 (2016).
11. M. JOCKENHÖVEL-BARTTFELD, A. TAURINES and C. HESSLER, Quantification of Application Software Failures of Digital I&C in Probabilistic Safety Analyses. *13th International Conference on Probabilistic Safety Assessment and Management*, Seoul, Korea (2016, October).
12. J. SHIN, H. SON and G. HEO, Development of A Cyber Security Risk Model Using Bayesian Networks. *Reliability Engineering & System Safety*, 134, 208-217 (2015).
13. S. KRIAA et al., A Survey of Approaches Combining Safety and Security for Industrial Control Systems. *Reliability Engineering & System Safety*, 139, 156-178 (2015).
14. W. WANG, F. DI MAIO and E. ZIO, Component-And System-Level Degradation Modeling of Digital Instrumentation and Control Systems Based on a Multi-State Physics Modeling Approach. *Annals of Nuclear Energy*, 95, 135-147 (2016).
15. R. C. MACHADO et al., Software Control and Intellectual Property Protection in Cyber-Physical Systems. *EURASIP Journal on Information Security*, 2016(1), 1-14 (2016).
16. O. BÄCKSTROM et al., Software Reliability Analysis For PSA: Failure Mode and Data, *Nordic nuclear safety research (NKS) Report*, NKS-341 (2015).
17. T. ALDEMIR et al., Probabilistic Risk Assessment Modeling of Digital Instrumentation and Control Systems Using Two Dynamic Methodologies. *Reliability Engineering & System Safety*, 95(10), 1011-1039 (2010).
18. F. KHORRAMI, P. KRISHNAMURTHY and R. KARRI, Cybersecurity for Control Systems: A Process-Aware Perspective. *IEEE Design & Test*, 33(5), 75-83 (2016).
19. Y. SOUPIONIS, S. NTALAMPIRAS and G. GIANNOPOULOS, Faults and Cyber Attacks Detection in Critical Infrastructures. *In International Conference on Critical Information Infrastructures Security*, p. 283-289. Springer International Publishing (2014, October).
20. S. A. EIDE et al., *Industry-Average Performance for Components and Initiating Events at US Commercial Nuclear Power Plants*. Idaho National Laboratory, US Nuclear Regulatory Commission Office of Nuclear Regulatory Research Washington, DC, 20555-0001 (2007).
21. E. ZIO, *An Introduction to The Basics of Reliability and Risk Analysis (Vol. 13)*. World scientific (2007).
22. US: EPRI. *Utility Requirement Document Annex a Reliability Data Base for Passive ALWR PRAs* (2008).