Modeling and Analysis of Attacks and Counter Defense Mechanisms for Cyber Physical Systems

Robert Mitchell and Ing-Ray Chen, *Member, IEEE*

***Abstract*—In this paper, we develop an analytical model based on stochastic Petri nets to capture the dynamics between adversary behavior and defense for cyber physical systems. We consider several types of failures including attrition failure, pervasion failure, and exfiltration failure which can happen to a cyber physical system. Using a modernized electrical grid as an example, we illustrate the parameterization process. Our results reveal optimal design conditions, including the intrusion detection interval, and the redundancy level, under which the modernized electrical grid's mean time to failure is maximized. Further, there exists a design tradeoff between exfiltration failure, attrition failure, and pervasion failure when using redundancy to improve the overall system reliability.**

***Index Terms*—Cyber physical systems, intrusion detection, redundancy engineering, mean time to failure, modeling and analysis.**

Acronyms and Abbreviations

CPS	Cyber physical system
IDS	Intrusion detection system
MTTF	Mean time to failure
SPN	Stochastic Petri net

Notation

S, C, A	Sensors, control nodes, actuators
INITx	Initial node type x count
MINx	Minimum node type x count
PBADx	Compromised node type x count
PGOODx	Uncompromised node type x count
$\mathrm{T}_{\mathrm{sensing}}$	Sensing interval
T_{TX}	Exfiltration rate
$\mathcal{P}_{\mathrm{fnx}}$	Node type x false negative probability
$\mathcal{P}_{\mathrm{fpx}}$	Node type x false positive probability
$\mathrm{T}_{\mathrm{IDSx}}$	Node type x audit interval
α_{x}	Node type x redundancy factor
λ_{x}	Node type x compromise rate
λ_{TCPx}	Node type x aggregate compromise rate
λ_{TFPx}	Node type x aggregate false positive rate
λ_{TIDx}	Node type x aggregate detection rate
$\lambda_{\mathrm{TLEAKx}}$	Node type x aggregate exfiltration rate

I. Introduction

While the importance of the survivability of cyber physical systems (CPSs) against malicious attacks is well recognized, the literature is thin in modeling and analysis of attacks and counter countermeasures for CPSs [1], [2]. To date, there are two lines of research in modeling and analysis of CPSs. The first line of work focused on a formal process or framework for designing and engineering a CPS [3]–[6]. The basic idea for this line of work is to formalize safety and functional requirements utilizing formal modeling and analysis tools, and then perform rigorous model verification. The second line of work focused on a mathematical model for analyzing the system's response behavior in the presence of malicious nodes performing various attacks [7]–[11]. The basic idea is to develop a state-based stochastic process to model a CPS equipped with an intrusion detection system (IDS) presented with various types of attacks, including random, opportunistic, and insidious, with the objective to improve IDS designs so as to prolong the system lifetime.

We follow the second line of research work with the primary objective to capture the dynamics between adversary behavior and defense for survivability of CPSs. The end product is a tool that is capable of analyzing a myriad of attacker behaviors, and seeing the effectiveness of countering adaptive defense strategies which incorporate attack-response dynamics. Relative to the works cited above, our contribution is threefold.

- First, we study the effect of attack and countermeasures on the survivability of CPSs. To the best of our knowledge, we are the first to develop an analytical model to capture the dynamics between adversary and defense for CPSs, as a result of applying attacks and countermeasures.
- Second, we define three failure types in CPSs, namely, attrition, pervasion, and exfiltration failure which can happen to a cyber physical system. Using a modernized electrical grid as an example, we illustrate the parameterization process. Our results reveal optimal design conditions, including the intrusion detection interval, and the redundancy level, under which the modernized electrical grid's mean time to failure (MTTF) is maximized.
- Third, our analytical model paves the way to answer what if questions. In this paper, we determine if using redundancy to cope with attacks for system survivability is viable in a CPS. Our result reveals that there exists a design tradeoff between exfiltration, attrition, and pervasion failure when using redundancy to improve the overall system reliability.

Manuscript received September 09, 2014; revised October 30, 2014; accepted February 22, 2015. Date of current version March 01, 2016. This work was supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under contract number W911NF-12-1-0445. Associate Editor: G Levitin.

The authors are with the Department of Computer Science, Virginia Polytechnic Institute and State University, Falls Church, VA 22043 USA (e-mail: rrmitche@vt.edu; irchen@vt.edu).

Digital Object Identifier 10.1109/TR.2015.2406860

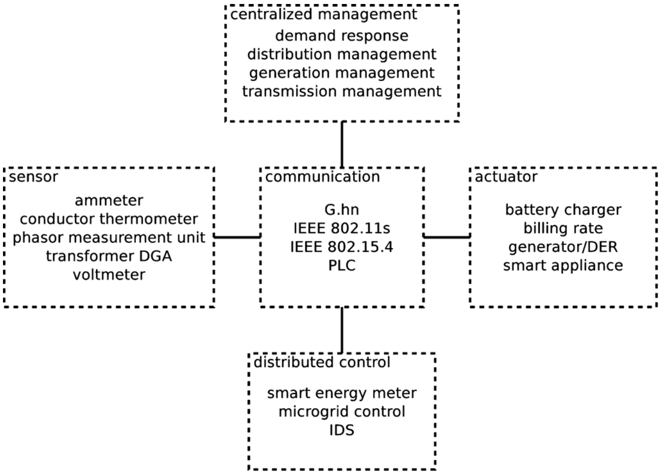

Fig. 1. Abstraction model for a modernized electrical grid.

The rest of the paper is organized as follows. In Section II, we introduce the system model including the abstraction model of a CPS, system failure definitions, attacker behaviors, and countermeasures considered in this paper. In Section III, we develop an analytical model based on stochastic Petri net (SPN) techniques [12]–[14] for modeling and analysis of attacks and counter countermeasures for CPSs. In Section IV, using a modernized electrical grid as an example, we illustrate the parameterization process, i.e., assigning model parameters with values, and present numerical results. Finally, in Section V, we conclude the paper, and outline future areas.

II. System Model

A. System Description

A modernized electrical grid is a smart grid that uses digital information and communications technology to gather and act on information, such as information about the behaviors of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity [15]. Our intention is to provide an abstraction model for a modernized electrical grid equipped with specific physical devices to illustrate the effect of attacks and countermeasures on system survivability.

Fig. 1 illustrates the abstraction model for a modernized electrical grid. For ease of disposition, this paper is particularly concerned with five types of physical devices: centralized management nodes, sensors, distributed control nodes, actuators, and communication links. This classification can frame most real and imagined attack scenarios. Centralized management nodes are attended, physically secure, and high-performance; they perform system-wide management functions. Sensors are unattended, physically vulnerable, and economical; they translate measurements of the physical world into the cyber domain. Distributed control nodes are also unattended, physically vulnerable, and economical. These nodes serve as agents for the centralized management nodes; they also execute control algorithms on sensor data, and apply the results to the actuators. Actuators are also unattended, physically vulnerable, and economical; they translate decisions made in the cyber domain into the physical world. Communication links connect centralized management nodes, sensors, control nodes, and actuators.

B. System Failure Definition

We consider three types of system failure.

- *Attrition failure* occurs when the modernized electrical grid doesn't have enough control nodes or actuators to accomplish its intended functions. Intuitively, no evicted or compromised nodes work toward the objective. Our model doesn't consider sensors in attrition failure. On one hand, if a sensor is evicted, the short-term impact is minimal: any control loop can run free of external input long enough to restore the evicted sensor. On the other hand, if a sensor is compromised, it can do little more than send illegitimate data to a control node where its dissenting voice will be drowned out by the preponderance of uncompromised sensors sending legitimate data.
- *Pervasion failure* occurs when the density of compromised control nodes or actuators is too high. In this situation, compromised nodes collude to overwhelm the other nodes. We don't consider sensors in pervasion failure, because a compromised sensor has no means to directly or indirectly attack the modernized electrical grid. In terms of Fig. 1, an adversary would prosecute pervasion failure via an uncompromised control node tasking an uncompromised actuator. An attack using two uncompromised, but adversary-compliant, nodes simultaneously is a fragile proposition, so our model doesn't include it.
- *Exfiltration failure* occurs when the aggressor secretes enough modernized electrical grid data to achieve an intelligence victory, or leaks enough surveillance data to instrument a devastating attack. Unlike the direct mission impact that attrition failure requires, and the direct means to damage the modernized electrical grid that pervasion failure requires, exfiltration is perfectly suited to compromised sensors because receiving raw data is a sensor's sole purpose. After gathering sensing reports, a compromised control node can also leak information. We consider sensors and control nodes in the exfiltration failure analysis. The basic sequence of events in an exfiltration attack is
  1) the aggressor is authenticated on the victim network,
  2) the aggressor finds valuable data,
  3) the aggressor connects with an aggressor-owned server outside of the victim network,
  4) the aggressor transmits the valuable data, and
  5) the victim experiences exfiltration failure.

C. Attacker Behavior Modeling

1) Surveilling Attacker: This brand of attacker seeks to gain information about or information residing on the target system. A surveillance aircraft is a kinetic warfare analog to this type of attacker. In the commercial domain, a company would do this to steal trade secrets from a competitor. This brand of attacker is interested in long-term operations so they will go to great lengths (even degrading their own mission) to avoid detection. In terms of Fig. 1, the surveilling attacker has more interest in centralized management nodes, communication links, and sensors; and less interest in actuators, and control nodes.

2) Destructive Attacker: This brand of attacker seeks to disrupt the target system. A bomb is a kinetic warfare analog to this case. In the law enforcement domain, a political group would do this to disrupt some entity with a different worldview. This brand of attacker is not concerned with discretion, and will act

TABLE I
COUNTERMEASURES, MEANINGS, AND DESIGN PARAMETERS

Mechanism	Physical meaning	Design parameters
Intrusion detection (anomaly and signature based)	The system performs an intrusion detection audit on a target node in every $\mathrm{T_{IDSx}}$ interval.	$\mathrm{T_{IDSx}}$
Data leak rate control	The system controls data leak rate by instrumenting a firewall to throttle outbound traffic ($\mathrm{T_{TX}}$) and rotating one sensor among all sensors measuring the same physical phenomenon to do sensing and data transmission per sensing interval ($\mathrm{T_{sensing}}$).	$\mathrm{T_{TX}}$, $\mathrm{T_{sensing}}$
Redundancy	The system installs more nodes than needed to prevent attrition failure.	α_{x} (Note: MINx is an input parameter and $\mathrm{INITx} = \mathrm{MINx} \times \alpha_{\mathrm{x}}$ where $\mathrm{x} \in \{\mathrm{C}, \mathrm{A}\}$.)

TABLE II
ATTACK BEHAVIOR, SYSTEM FAILURE TYPE, COUNTERMEASURE AND VULNERABLE DEVICES

Attack behavior	System failure type	Countermeasure	Vulnerable devices
Surveilling attacker	Exfiltration failure	Data leak rate control	Sensors and control nodes
Destructive attacker	Attrition failure	Redundancy	Distributed control nodes and actuators
Destructive attacker	Pervasion failure	Intrusion detection	Distributed control nodes and actuators

with impunity. In terms of Fig. 1, the destructive attacker has more interest in actuators, centralized management nodes, and control nodes; and less interest in communication links, and sensors. One way to disrupt the system is to reduce the number of control nodes and actuators operating correctly. Another way to disrupt the system is to pervade control nodes and actuators discreetly. In these scenarios, the centralized management nodes are not likely targets because they are physically controlled, highly reliable, and under close scrutiny from computer security software.

TABLE III
PLACES IN THE SPN PERFORMANCE MODEL

Place	Meaning
PATTRIT	0 before attrition failure, and 1 after
PGOODS	Unevicted, uncompromised sensor count
PGOODC	Unevicted, uncompromised control node count
PGOODA	Unevicted, uncompromised actuator count
PBADS	Unevicted, compromised sensor count
PBADC	Unevicted, compromised control node count
PBADA	Unevicted, compromised actuator count
PLEAK	0 before exfiltration failure, and 1 after
PPERVADE	0 before pervasion failure, and 1 after

D. Countermeasures

1) Intrusion Detection: The CPS has an IDS applying anomaly or signature based detection techniques [1], [16], [17] to detect and evict suspicious nodes. The intrusion detection quality is characterized by the false negative probability ($\mathcal{P}_{\mathrm{fnx}}$), and false positive probability ($\mathcal{P}_{\mathrm{fpx}}$) with $\mathrm{x} \in \{\mathrm{S}, \mathrm{C}, \mathrm{A}\}$ for sensors, control nodes, and actuators, respectively. The former quality metric defines the probability that a malicious node is misdetected, while the latter quality metric defines the probability that a good node is misidentified as a malicious node. We assume that $\mathcal{P}_{\mathrm{fnx}}$, and $\mathcal{P}_{\mathrm{fpx}}$ with $\mathrm{x} \in \{\mathrm{S}, \mathrm{C}, \mathrm{A}\}$ are inputs to our model.

The countermeasure employed by the CPS to detect and evict malicious devices is to apply the optimal detection interval $\mathrm{T_{IDSx}}$ for periodic intrusion detection with $\mathrm{x} \in \{\mathrm{S}, \mathrm{C}, \mathrm{A}\}$, for sensors, control nodes, and actuators, respectively. When $\mathcal{P}_{\mathrm{fnx}}$ is low, the system can benefit from a small intrusion detection interval because malicious nodes can be detected and evicted often. On the other hand, when $\mathcal{P}_{\mathrm{fpx}}$ is high, the system can benefit from a large intrusion detection interval because good nodes won't be misidentified as malicious nodes, and mistakenly evicted often. Hence, identifying and applying the optimal detection interval $\mathrm{T_{IDSx}}$ to best balance $\mathcal{P}_{\mathrm{fnx}}$ and $\mathcal{P}_{\mathrm{fpx}}$ can enhance the system MTTF.

2) Data Leak Rate Control: The CPS prevents or delays exfiltration failure by data leak rate control. To cope with compromised sensors and control nodes, it runs an inward facing firewall. In the scenario described in Section II.B, when the aggressor attempts to connect to a server outside of the network, the firewall may deny the connection because it is not authorized, or the IP address of the server is not on a whitelist. Even if these rules are not active on the firewall, it may throttle the outbound session (e.g., 100 kbps). If the valuable dataset is, for example, a 1 GB dataset, this would buy the victim almost 24 hours to detect the leak and evict the aggressor. The critical design parameter of this countermeasure is a maximum transmission rate of $\mathrm{T_{TX}}$ bits per second.

To cope with compromised sensor nodes leaking sensing results, the system limits the data leak rate by rotating one sensor among all sensors measuring the same physical phenomenon, to do sensing and data transmission per sensing interval ($\mathrm{T_{sensing}}$). The critical design parameter of this countermeasure is $\mathrm{T_{sensing}}$, over which data leakage is possible only when the compromised sensor node is rotated in to do sensing. If a sensor performs data transmission in every $\mathrm{T_{sensing}}$ interval, the IDS generates a detection.

3) Redundancy: Modern electrical grid systems use some degree of *redundancy* to counterbalance failed, evicted, and compromised nodes. The critical design parameter of this countermeasure is the redundancy factor (α_{x}) over the minimum number of nodes (MINx) required for functionality such that the number of nodes initially put into the system is

$$\mathrm{INITx} = \mathrm{MINx} \times \alpha_{\mathrm{x}} \tag{1}$$

where $\mathrm{x} \in \{\mathrm{C}, \mathrm{A}\}$. We aim to analyze the design parameter settings for performance maximization in terms of the MTTF. Table I provides additional detail on the countermeasures we consider.

Table II summarizes the relationship between the attack behavior, countermeasure, and malfunctioned device type causing a system failure.

TABLE IV
TRANSITIONS IN THE SPN PERFORMANCE MODEL

Transition	Meaning
TFPS	IDS falsely detects a sensor
TFPC	IDS falsely detects a control node
TFPA	IDS falsely detects an actuator
TCPS	Attacker compromises a sensor
TCPC	Attacker compromises a control node
TCPA	Attacker compromises an actuator
TIDS	IDS detects a compromised sensor
TIDC	IDS detects a compromised control node
TIDA	IDS detects a compromised actuator
TLEAKS	Attacker secretes a substantial amount of victim sensor data
TLEAKC	Attacker secretes a substantial amount of victim control node data
TATTRITC, TATTRITA	Immediate transition governed by EATTRITx
TPERVADEC, TPERVADEA	Immediate transition governed by EPERVADEx

TABLE V
ENABLING FUNCTIONS IN THE SPN PERFORMANCE MODEL

Function name	Enabling condition
EATTRITC	Uncompromised control node count is less than the minimum count
EATTRITA	Uncompromised actuator count is less than the minimum count
EPERVADEC	Byzantine failure condition applied to control nodes
EPERVADEA	Byzantine failure condition applied to actuators

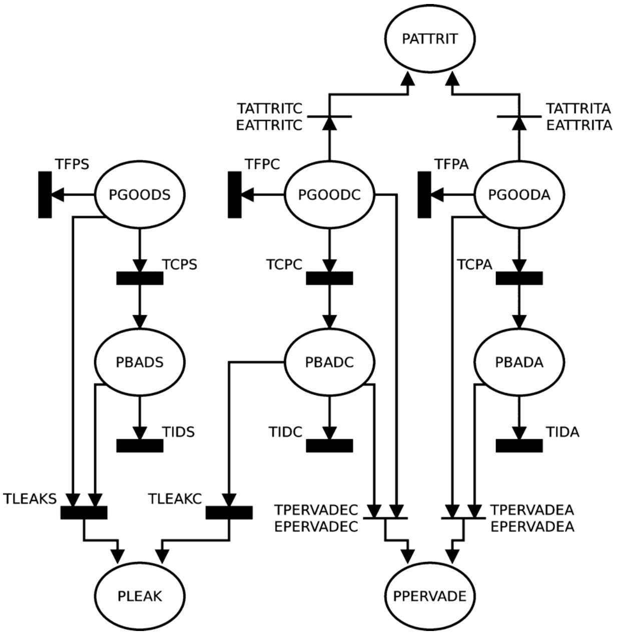

Fig. 2. SPN model.

III. PERFORMANCE MODEL

In this section, we develop a performance model as shown in Fig. 2 based on SPN modeling techniques to describe the system behavior in the presence of attacker behavior and countermeasures. Tables III and IV annotate the physical meanings of places and transitions in the SPN model. Table V defines the enabling functions for firing transitions in the SPN model. For simplicity, we consider three devices: sensors, control nodes, and actuators.

The underlying semi-Markov model has a 9-state representation: (PATTRIT, PGOODS, PGOODC, PGOODA, PBADS, PBADC, PBADA, PLEAK, PPERVADE). The underlying model would be Markov if transition times were exponentially distributed. However, this is a strong assumption, so we use a semi-Markov model to underlie the SPN to accommodate generally distributed transition times.

The underlying semi-Markov model has nine places. The PATTRIT place, if holding a token, represents a system failure resulting from too many control nodes or actuators being evicted or compromised. The PGOODS, PGOODC, and PGOODA places hold the count of un-evicted and uncompromised sensors, control nodes, and actuators, respectively. Similarly, the PBADS, PBADC, and PBADA places hold the count of un-evicted and compromised sensors, control nodes, and actuators, respectively. The PPERVADE place, if holding a token, represents a system failure resulting from a high ratio of compromised to uncompromised control nodes or actuators. The PLEAK place, if holding a token, represents a system failure resulting from compromised sensors and control nodes exfiltrating too much data.

The SPN model is constructed as follows.

- The first event is the system initialization by which we populate the system with INITx nodes with $x \in \{S, C, A\}$, for sensors, control nodes, and actuators, respectively. We use places to hold tokens with each token representing one node. Initially, all nodes are uncompromised, and put in places PGOODx as tokens.
- The next event we consider is the adversary compromising an uncompromised node. Transitions TCPx model this event with $x \in \{S, C, A\}$, for sensors, control nodes, and actuators, respectively. Uncompromised nodes may become compromised because of capture events. We assume that the time for the adversary to capture a node of type x (which may be a sensor, a control node, or an actuator) and convert it into a malicious node is a random variable following a distribution function (e.g., an exponential distribution). This event is modeled by associating transitions TCPx with rates λ_{TCPx}. Firing TCPx will move tokens (if available) one at a time from place PGOODx to place PBADx. Tokens in place PBADx represent unevicted compromised nodes. Fig. 3 illustrates

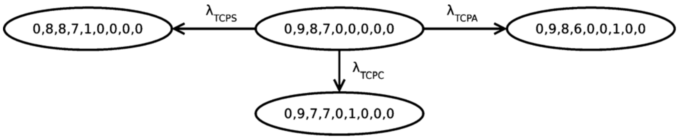

Fig. 3. Node captures in the underlying semi-Markov model.

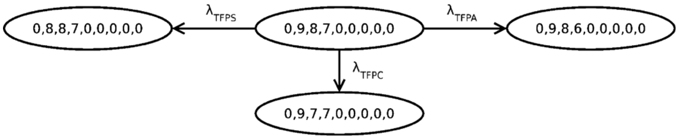

Fig. 4. False positives in the underlying semi-Markov model.

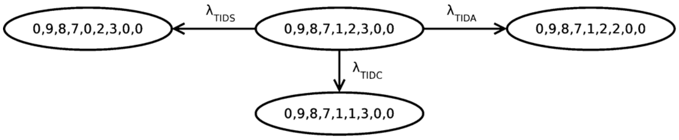

Fig. 5. Detections in the underlying semi-Markov model.

these transitions. For example, if in state (0, n_s, n_c, n_a, 0, 0, 0, 0, 0), an uncompromised sensor node is compromised, a token will flow from PGOODS to PBADS, and the resulting state is (0, $n_s - 1$, n_c, n_a, 1, 0, 0, 0, 0).

- The third event we consider is the IDS incorrectly evicting an uncompromised node. Transitions TFPx model this event with x $\in \{\mathrm{S}, \mathrm{C}, \mathrm{A}\}$, for sensors, control nodes, and actuators, respectively. Uncompromised nodes may be evicted because of intrusion detection error. This event is modeled by removing an uncompromised node from place PGOODx by firing transitions TFPx with rates of λ_{TFPx}. Fig. 4 illustrates these transitions. For example, if in state (0, n_s, n_c, n_a, 0, 0, 0, 0, 0) the IDS misdetects and evicts an uncompromised actuator, a token will flow from PGOODA, and the resulting state is (0, n_s, n_c, $n_a - 1$, 0, 0, 0, 0, 0).
- The next event we consider is the IDS correctly evicting a compromised node. Transitions TIDx model this event with x $\in \{\mathrm{S}, \mathrm{C}, \mathrm{A}\}$, for sensors, control nodes, and actuators, respectively. When a compromised node is detected as compromised, the number of unevicted compromised nodes will be decremented by one, i.e., place PBADx will hold one less token. This event is modeled by associating transitions TIDx with rates λ_{TIDx}. Fig. 5 illustrates these transitions. For example, if in state (0, n_s, $n_c - 1$, n_a, 0, 1, 0, 0, 0) the IDS detects and evicts a compromised control node, a token will flow from PBADC, and the resulting state is (0, n_s, n_c − 1, n_a, 0, 0, 0, 0, 0). The physical meaning of the TIDx timed transitions is the rate that the modernized electrical grid IDS generates true positives for compromised sensors, control nodes, and actuators.
- The fifth event we consider is the system failing due to attrition. That is, the system fails when the number of nodes with node type x is less than the minimum specified by MINx. TATTRITx models this attrition failure event with x $\in \{\mathrm{C}, \mathrm{A}\}$, for control nodes, and actuators, respectively. Table V lists the enabling functions governing the firing of TATTRITx. When TATTRITx is enabled, that is, the attrition failure condition is true, then the corresponding enabling function returns true. This condition will put a token into place PATTRIT, representing that an attrition failure has occurred. Physically, the transition TATTRITx is enabled and fired when the number of nodes is less than MINx. Fig. 6 illustrates this event from the perspective of the underlying semi-Markov chain.

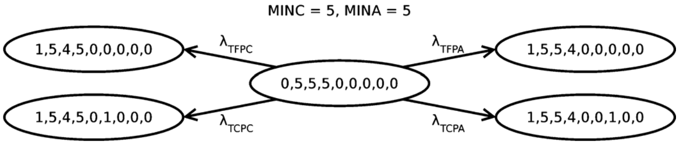

Fig. 6. Attrition failure in the underlying semi-Markov model.

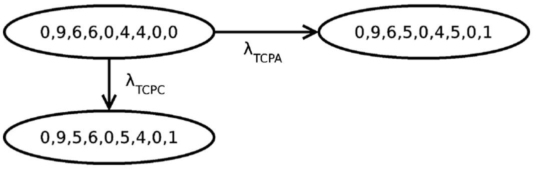

Fig. 7. Pervasion failure in the underlying semi-Markov model.

- The next event we consider is the system failing due to pervasion. TPERVADEx models this pervasion failure event, with x $\in \{\mathrm{C}, \mathrm{A}\}$, for control nodes, and actuators, respectively. When uncompromised control nodes and actuators introduced in Fig. 1 transition to compromised (PBADx), they degrade the defense of the network by falsely endorsing their confederates, and falsely reporting uncompromised nodes as compromised. Also, when the modernized electrical grid evicts uncompromised nodes (TFPx), this reduces the preponderance of uncompromised nodes counterbalancing the false endorsements and false alerts. This defense can be defeated when at least 1/3 of the control nodes or actuators introduced in Fig. 1 are compromised (PBADx) following the definition of Byzantine failure [18]. The enabling functions of TPERVADEx with x $\in \{\mathrm{C}, \mathrm{A}\}$ are defined in Table V governing the firing of TPERVADEx. When TPERVADEx is enabled, that is, the pervasion failure condition is true, then the corresponding enabling function returns true. This action will put a token into place PPERVADE, representing that a pervasion failure has occurred. Fig. 7 illustrates this event from the perspective of the underlying semi-Markov chain.
- The seventh event we consider is the system failing due to extensive exfiltration. TLEAKx models this failure event, with x $\in \{\mathrm{S}, \mathrm{C}\}$ for sensors, and control nodes, respectively. When compromised sensor nodes (PBADS) discreetly transmit or compromised control nodes (PBADC) discreetly relay the confidential data of a modernized electrical grid outside the system, competitors and criminals learn valuable business intelligence, and guerrillas and nation-states learn of system vulnerabilities. Data leak rate controls (i.e., $\mathrm{T_{TX}}$ and $\mathrm{T_{sensing}}$) are our countermeasures for this threat. This defense can be defeated given enough time for data exfiltration. The physical meaning of the TLEAKx transition is the event that the aggressor secretes enough data to cause an exfiltration failure. Fig. 8 illustrates this event from the perspective of the underlying semi-Markov chain.

TABLE VI
INPUT AND DESIGN PARAMETERS AND THEIR VALUES FOR A MODERNIZED ELECTRICAL GRID

Parameter name	Physical meaning	Default value	Type
$\mathrm{T_{IDSx}}$	Intrusion detection interval for node type $\mathrm{x} \in \{\mathrm{S, C, A}\}$	variable	Design
$\mathrm{T_{TX}}$	Maximum upload rate	(0, 1] Mbps	Design
$\mathrm{T_{sensing}}$	Sensing interval	variable	Design
α_{x}	Node redundancy factor for node type $\mathrm{x} \in \{\mathrm{C, A}\}$	1.5, 3, 4.5	Design
λ_{x}	Per-node compromise rate for node type $\mathrm{x} \in \{\mathrm{S, C, A}\}$	variable	Input
MAXLEAKS	Maximum number of sensor reading leaks	variable	Input
MAXLEAKC	Maximum amount of data leaked	variable	Input
MINC	Minimum number of control nodes	2	Input
MINA	Minimum number of actuators	4	Input
INITS	Initial number of sensor nodes	32	Input
INITC	Initial number of control nodes	MINC $\times\ \alpha_{\mathrm{C}}$	Input
INITA	Initial number of actuator nodes	MINA $\times\ \alpha_{\mathrm{A}}$	Input
$\mathcal{P}_{\mathrm{fpx}}$	IDS false positive probability for node type $\mathrm{x} \in \{\mathrm{S, C, A}\}$	$0.1 - 0.5$	Input
$\mathcal{P}_{\mathrm{fnx}}$	IDS false negative probability for node type $\mathrm{x} \in \{\mathrm{S, C, A}\}$	$0.1 - 0.5$	Input

TABLE VII
PARAMETERIZATION OF TRANSITION RATES IN THE SPN MODEL

Parameter name	Physical meaning	Parameterization
λ_{TCPx}	Aggregate compromise rate	$\mathrm{\|PGOODx\|} \times \lambda_{\mathrm{x}}$
λ_{TIDx}	Aggregate detection rate	$\mathrm{\|PBADx\|} \times (1 - \mathcal{P}_{\mathrm{fnx}})/\mathrm{T_{IDSx}}$
λ_{TFPx}	Aggregate false positive rate	$\mathrm{\|PGOODx\|} \times \mathcal{P}_{\mathrm{fpx}}/\mathrm{T_{IDSx}}$
$\lambda_{\mathrm{TLEAKS}}$	Aggregate sensor exfiltration rate	$\mathrm{\|PBADS\|/(\|PBADS\| + \|PGOODS\|)} \times 1/\mathrm{T_{sensing}} \times 1/\mathrm{MAXLEAKS}$
$\lambda_{\mathrm{TLEAKC}}$	Aggregate control node exfiltration rate	$\mathrm{\|PBADC\|} \times \mathrm{T_{TX}} \times 1/\mathrm{MAXLEAKC}$

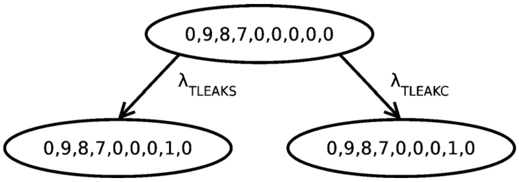

Fig. 8. Exfiltration failure in the underlying semi-Markov model.

IV. PERFORMANCE ANALYSIS

A. Model Parameterization

Table VI lists the input parameters and their default values or ranges of values used for the modernized electrical grid described in our system model. A design parameter is one that the system manager can choose. On the other hand, an input parameter is one that the operating environment dictates.

We perform the parameterization processes (i.e., give values to model parameters) for the transition rates in the SPN model. Let λ_{T} be the transition rate of transition T in the SPN model. Table VII summarizes the parameterization of λ_{TCPx}, λ_{TIDx}, λ_{TFPx}, and $\lambda_{\mathrm{TLEAKx}}$ from the input and design parameters listed in Table VI. Below we provide physical explanations.

λ_{TCPx} can be derived using the formulation

$$\lambda_{\mathrm{TCPx}} = |\mathrm{PGOODx}| \times \lambda_{\mathrm{x}} \tag{2}$$

where $|\mathrm{PGOODx}|$ is the number of uncompromised nodes of device type x, and λ_{x} is the per-node compromise rate for device type x. Intuitively, more uncompromised sensors, control nodes, or actuators translates to more opportunities for compromise.

λ_{TIDx} can be derived using the formulation

$$\lambda_{\mathrm{TIDx}} = |\mathrm{PBADx}| \times (1 - \mathcal{P}_{\mathrm{fnx}})/\mathrm{T_{IDSx}} \tag{3}$$

where $|\mathrm{PBADx}|$ is the number of compromised nodes, $\mathcal{P}_{\mathrm{fnx}}$ is the false negative probability, and $\mathrm{T_{IDSx}}$ is the IDS detection interval for device type x. Intuitively, in every $\mathrm{T_{IDSx}}$ interval, a bad node of type x will be correctly identified as a bad node with probability $1 - \mathcal{P}_{\mathrm{fnx}}$, so the aggregate rate at which bad nodes are detected and evicted correctly is $|\mathrm{PBADx}|$ multiplied with $(1 - \mathcal{P}_{\mathrm{fnx}})/\mathrm{T_{IDSx}}$.

λ_{TFPx} can be derived using the formulation

$$\lambda_{\mathrm{TFPx}} = |\mathrm{PGOODx}| \times \mathcal{P}_{\mathrm{fpx}}/\mathrm{T_{IDSx}} \tag{4}$$

where $|\mathrm{PGOODx}|$ is the number of uncompromised nodes, $\mathcal{P}_{\mathrm{fpx}}$ is the false positive probability, and $\mathrm{T_{IDSx}}$ is the IDS detection interval for device type x. Intuitively, in every $\mathrm{T_{IDSx}}$ interval, a good node of type x will be misidentified as a bad node with probability $\mathcal{P}_{\mathrm{fpx}}$, so the aggregate rate at which good nodes suffer from false positives is $|\mathrm{PGOODx}|$ multiplied with $\mathcal{P}_{\mathrm{fpx}}/\mathrm{T_{IDSx}}$.

$\lambda_{\mathrm{TLEAKS}}$ can be derived using the formulation

$$\lambda_{\mathrm{TLEAKS}} = \frac{|\mathrm{PBADS}|}{|\mathrm{PBADS}| + |\mathrm{PGOODS}|} \times \frac{1}{\mathrm{T_{sensing}}} \times \frac{1}{\mathrm{MAXLEAKS}} \tag{5}$$

where the first term is for a compromised sensor node to rotate in for reporting sensing data, the second term is for the rate at which sensing reporting occurs, and the third term is for the maximum number of leaks (so MAXLEAKS is an input parameter) the system can tolerate before an exfiltration failure occurs. Note that, in the above formulation, MAXLEAKS is an input parameter.

$\lambda_{\mathrm{TLEAKC}}$ can be derived using the formulation

$$\lambda_{\mathrm{TLEAKC}} = |\mathrm{PBADC}| \times \mathrm{T_{TX}} \times 1/\mathrm{MAXLEAKC} \tag{6}$$

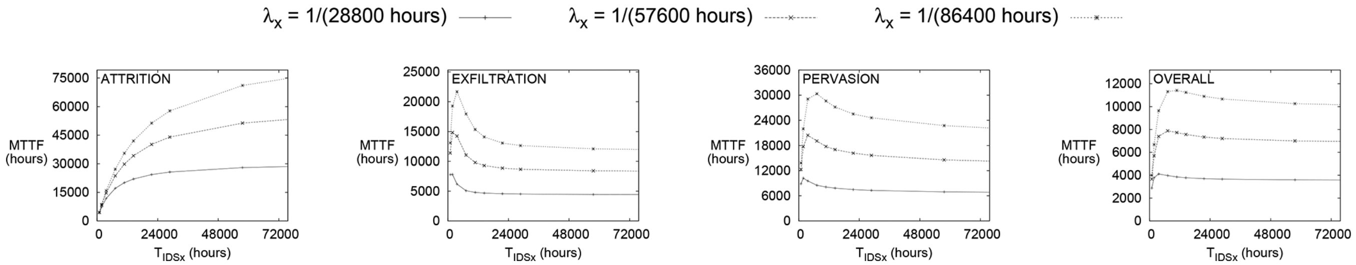

Fig. 9. MTTF vs. $\mathrm{T_{IDSx}}$ with varying $\lambda_{\mathrm{x}}(\mathcal{P}_{\mathrm{fn}} = 0.1, \mathcal{P}_{\mathrm{fp}} = 0.2, \alpha = 3)$.

where $\mathrm{T_{TX}}$ is the data transmission rate per node allowable, and MAXLEAKC is the maximum data amount leaked beyond which an exfiltration failure occurs. Note that, in the above formulation, MAXLEAKC is an input parameter.

B. Results

In this section, we present numerical data for MTTF assessment as a result of applying countermeasures described in Section II-D against attack behavior described in Section II-C causing attrition, pervasion, or exfiltration system failure. Our objective is to analyze the effect of countermeasures in terms of $\mathrm{T_{IDSx}}$, $\mathcal{P}_{\mathrm{fp}}$, $\mathcal{P}_{\mathrm{fn}}$, and α_{x} on MTTF, when given a set of parameter values characterizing the operational and networking conditions.

Let L be a binary random variable denoting the lifetime of the system such that it takes on the value of 1 if the system is alive at time t, and 0 otherwise. Then, the expected value of L is the reliability of the system R(t) at time t. Consequently, the integration of R(t) from t $=$ 0 to 1 gives the MTTF or the average lifetime of the system we aim to maximize. The binary value assignment to L can be done by means of a reward function assigning a reward r_i of 0 or 1 to state i at time t as

$$r_i = \begin{cases} 1 & \text{if the system is alive in state } i, \\ 0 & \text{if the system fails in state } i. \end{cases}$$

The MTTF computation methodology defined above requires the probability of the system being in state i at time t, $P_i(t)$, be known. This requirement is obtained by defining a SPN model using SPNP [19], and then solving the underlying semi-Markov model utilizing solution techniques such as SOR, Gauss Seidel, or Uniformization [19].

1) Effect of a Detection Interval $\mathrm{T_{IDSx}}$*:* Fig. 9 shows MTTF versus $\mathrm{T_{IDS}}$ (assuming $\mathrm{T_{IDSS}} = \mathrm{T_{IDSC}} = \mathrm{T_{IDSA}}$) with varying attack intensity λ_{x}. We set $\mathcal{P}_{\mathrm{fn}} = 0.1$, $\mathcal{P}_{\mathrm{fp}} = 0.2$, and $\alpha = 3$ to isolate out their effects. From left to right, the four graphs are for the cases of attrition failure only, exfiltration failure only, pervasion failure only, and combined failure, respectively. We first observe that MTTF decreases as the attacker strength increases, as expected, for all failure types. An important observation is that, except for attrition failure, there is an optimal $\mathrm{T_{IDS}}$ value under which the MTTF is maximized. We explain the reason as follows.

- The reason that the MTTF under attrition failure increases monotonically as $\mathrm{T_{IDS}}$ increases (in the leftmost graph) is due to the setting of $\mathcal{P}_{\mathrm{fn}} = 0.1 < \mathcal{P}_{\mathrm{fp}} = 0.2$. That is, the probability that a good node is misidentified as a bad node is higher than the probability that a bad node is missed. Consequently, a higher frequency of intrusion detection, i.e., a smaller $\mathrm{T_{IDS}}$, will cause more good nodes to be evicted than bad nodes evicted, thus causing the system to fail faster due to attrition failure because of a lack of good nodes in the system.
- The reason that the MTTF under exfiltration failure is maximized at the optimal $\mathrm{T_{IDS}}$ value identified (in the second leftmost graph) is due to the fact that the exfiltration failure rate is mainly affected by the bad node ratio, i.e., the ratio of the number of bad nodes to the total number of bad and good nodes, as explained in Table VII. Therefore, to maximize the MTTF under exfiltration failure, one needs to minimize this ratio. The optimal $\mathrm{T_{IDS}}$ that maximizes the MTTF under exfiltration failure exists because the bad node ratio minimizes with this optimal $\mathrm{T_{IDS}}$ value.
- The reason that the MTTF under pervasion failure is maximized at the optimal $\mathrm{T_{IDS}}$ value identified (in the second rightmost graph) is due to the fact that pervasion failure occurs when the bad node ratio is at least 1/3. Therefore, to prevent pervasion failure, the bad node ratio must be kept below 1/3. The optimal $\mathrm{T_{IDS}}$ value that maximizes the MTTF under pervasion failure exists because, with this optimal $\mathrm{T_{IDS}}$ value, the bad node ratio is the lowest.

The rightmost graph shows the system MTTF versus $\mathrm{T_{IDS}}$ when all failure types are considered. The MTTF curve in the rightmost graph essentially combines the three MTTF curves to the left. An important observation is that there still exists an optimal $\mathrm{T_{IDS}}$ for the MTTF curve under combined failure. Our model allows such an optimal $\mathrm{T_{IDS}}$ value to be identified when given a set of parameter values characterizing the operational and networking conditions.

2) Effect of False Positive Probability $\mathcal{P}_{\mathrm{fp}}$*:* Fig. 10 shows MTTF versus $\mathrm{T_{IDS}}$ (assuming $\mathrm{T_{IDSS}} = \mathrm{T_{IDSC}} = \mathrm{T_{IDSA}}$) with varying false positive probability $\mathcal{P}_{\mathrm{fp}}$. We set $\mathcal{P}_{\mathrm{fn}} = 0.1$, $\lambda_{\mathrm{x}} = 1/(2400 \text{ days})$, and $\alpha = 3$ to isolate their effects. From left to right, the four graphs are again for the cases of attrition failure only, exfiltration failure only, pervasion failure only, and combined failure, respectively. We first observe that MTTF decreases as $\mathcal{P}_{\mathrm{fp}}$ increases for all failure types because, as $\mathcal{P}_{\mathrm{fp}}$ increases, there is a higher probability of a good node being misidentified as a bad node and evicted. We observe the same trend as before. That is, except for attrition failure, there is an optimal $\mathrm{T_{IDS}}$ value under which the MTTF is maximized. The same physical explanations can be applied here. We also observe that the optimal $\mathrm{T_{IDS}}$ value for MTTF maximization increases as $\mathcal{P}_{\mathrm{fp}}$ increases. This result happens because, as $\mathcal{P}_{\mathrm{fp}}$ increases its magnitude relative to $\mathcal{P}_{\mathrm{fn}}$, intrusion detection may be detrimental to system reliability if it is performed too often because the rate at which good nodes are misidentified as bad nodes and evicted will increase relative to the rate at which bad nodes are detected and evicted. Consequently, as $\mathcal{P}_{\mathrm{fp}}$ increases

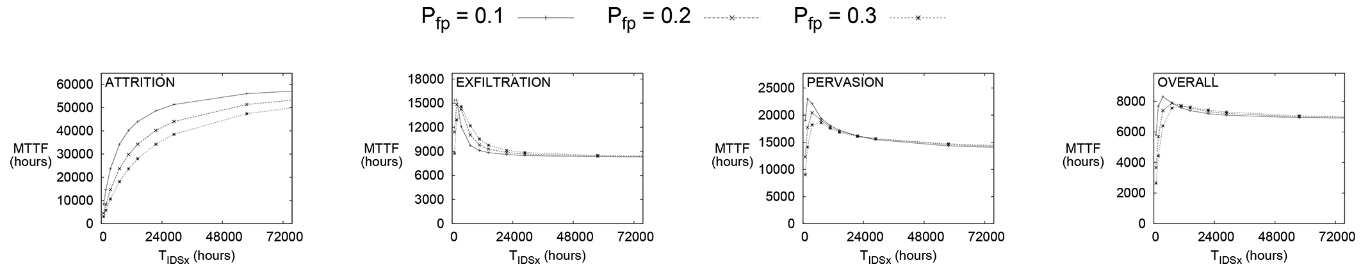

Fig. 10. MTTF vs. $\mathrm{T_{IDSx}}$ with varying $\mathcal{P}_{\mathrm{fp}}(\mathcal{P}_{\mathrm{fn}} = 0.1, \lambda_{\mathrm{x}} = 1/(2400\mathrm{days}), \alpha = 3)$.

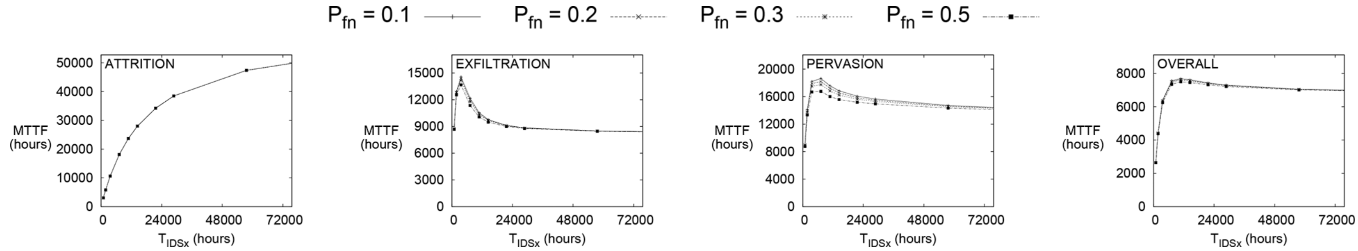

Fig. 11. MTTF vs. $\mathrm{T_{IDSx}}$ with varying $\mathcal{P}_{\mathrm{fn}}(\mathcal{P}_{\mathrm{fp}} = 0.3, \lambda_{\mathrm{x}} = 1/(2400\ \mathrm{days}), \alpha = 3)$.

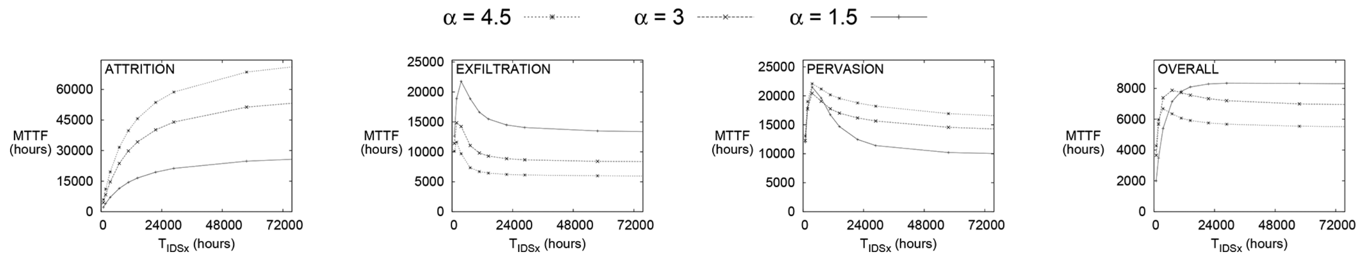

Fig. 12. MTTF vs. $\mathrm{T_{IDSx}}$ with varying $\alpha(\mathcal{P}_{\mathrm{fn}} = 0.1, \mathcal{P}_{\mathrm{fp}} = 0.2, \lambda_{\mathrm{x}} = 1/(2400\ \mathrm{days}))$.

its magnitude relative to $\mathcal{P}_{\mathrm{fn}}$, the optimal $\mathrm{T_{IDS}}$ value increases so as to minimize the bad node ratio. The rightmost graph of Fig. 10 shows the system MTTF versus $\mathrm{T_{IDS}}$ when all failure types are considered. We again observe that there still exists an optimal $\mathrm{T_{IDS}}$ for the MTTF curve under combined failure.

3) Effect of False Negative Probability $\mathcal{P}_{\mathrm{fn}}$: Fig. 11 shows MTTF versus $\mathrm{T_{IDS}}$ (assuming $\mathrm{T_{IDSS}} = \mathrm{T_{IDSC}} = \mathrm{T_{IDSA}}$) with varying false negative probability $\mathcal{P}_{\mathrm{fn}}$. We set $\mathcal{P}_{\mathrm{fp}} = 0.3$, $\lambda_{\mathrm{x}} = 1/(2400\ \mathrm{days})$, and $\alpha = 3$ to isolate their effects. From left to right, the four graphs are again for the cases of attrition failure only, exfiltration failure only, pervasion failure only, and combined failure, respectively. The trend exhibited in Fig. 11 for the effect of $\mathcal{P}_{\mathrm{fn}}$ is remarkably similar to that in Fig. 10 for the effect of $\mathcal{P}_{\mathrm{fp}}$ except that the MTTF is less sensitive to $\mathcal{P}_{\mathrm{fn}}$. In particular, the MTTF under attrition failure (the leftmost graph) is insensitive to $\mathcal{P}_{\mathrm{fn}}$. The reason is that attrition failure depends on the number of good nodes remaining in the system. Hence, attrition failure is only sensitive to the good node compromising rate, i.e., λ_{x}, which determines how fast a good node is compromised into a bad node, as well as the false positive rate, i.e., $\mathcal{P}_{\mathrm{fp}}$, which determines how fast a good node is misidentified as a bad node and evicted. The rightmost graph of Fig. 11 shows the system MTTF versus $\mathrm{T_{IDS}}$ when all failure types are considered. We again observe that there exists an optimal $\mathrm{T_{IDS}}$ for the MTTF curve under combined failure.

4) Effect of Redundancy Factor α_{x}: Fig. 12 shows MTTF versus $\mathrm{T_{IDS}}$ with varying redundancy α. We set $\mathcal{P}_{\mathrm{fp}} = 0.2$, $\mathcal{P}_{\mathrm{fn}} = 0.1$, and $\lambda_{\mathrm{x}} = 1/(2400\ \mathrm{days})$ to isolate their effects. From left to right, the four graphs are again for the cases of attrition failure only, exfiltration failure only, pervasion failure only, and combined failure, respectively. The redundancy factor α determines the number of nodes initially (INITx) with INITx $=$ MINx $\times\ \alpha_{\mathrm{x}}$ (where $\mathrm{x} \in \{\mathrm{C}, \mathrm{A}\}$), and MINx is the minimum number of control nodes or actuators, respectively, required to prevent attrition failure. Because attrition failure depends on the number of good nodes remaining in the system, putting in more initial nodes can better prevent attrition failure from occurring. Therefore, the MTTF under attrition failure (the leftmost graph) increases as α increases. It is interesting to observe from Fig. 12 that the MTTF under exfiltration failure (the second leftmost graph) decreases as α increases. This rather counter-intuitive result is due to the nature of exfiltration failure by control nodes or sensors. Specifically, there are two ways by which exfiltration failure can occur. One way is through TLEAKC which depends on the absolute number of bad control nodes (see Table VII for the TLEAKC rate $\lambda_{\mathrm{TLEAKC}}$). The other way is through TLEAKS which depends on the bad node ratio of sensors (also see Table VII for the TLEAKS rate $\lambda_{\mathrm{TLEAKS}}$). Among these two rates, $\lambda_{\mathrm{TLEAKC}}$ increases as the initial number of control nodes (that is, INITC) increases, i.e., as α_{C} increases, because this increases the chance of bad control nodes being produced due to node compromising events. The other rate, $\lambda_{\mathrm{TLEAKS}}$, increases as the bad node ratio of sensor nodes increases, which does not depend on α_{C}. Consequently, the MTTF under exfiltration failure (the second leftmost graph) decreases as α increases. We also observe that the MTTF under pervasion failure (the second rightmost graph) increases as α increases. This event happens because pervasion failure depends on the bad node ratio which decreases as more initial nodes are put in the system, especially if the detection interval $\mathrm{T_{IDS}}$ is large. Finally, the overall system MTTF (the rightmost graph) curve combines all MTTF curves to the left. We again observe that there exists an optimal $\mathrm{T_{IDS}}$ that maximizes the MTTF of the CPS against all attacks causing attrition, pervasion, or exfiltration system failures.

V. Conclusion

In this paper, we developed an analytical model based on SPNs to capture the dynamics between adversary behavior and defense for CPSs. Using a modernized electrical grid as an example, our results revealed optimal design conditions, including the intrusion detection interval, and the redundancy level under which the modernized electrical grid's MTTF is maximized. Further, we discovered that redundancy should be used with caution, because while it suppresses attrition and pervasion failure, it also induces exfiltration failure. The analytical model developed in this paper allows the optimal design parameter settings for maximizing the MTTF of the CPS to be identified, when given a set of input parameter values characterizing the operational and environment conditions.

In the future, we plan to investigate how control theory or game theory principles [20]–[22] controlling the attack-defense dynamics can further improve the CPS survivability.

References

[1] R. Mitchell and I. R. Chen, "A survey of intrusion detection techniques in cyber physical systems," *ACM Comput. Surveys*, vol. 46, no. 4, 2014.

[2] R. Mitchell and I. R. Chen, "A survey of intrusion detection in wireless network applications," *Comput. Commun.*, vol. 42, pp. 1–23, 2014.

[3] A. Banerjee, S. K. S. Gupta, G. Fainekos, and G. Varsamopoulos, "Towards modeling and analysis of cyber-physical medical systems," in *Proc. 4th Int. Symp. Applied Sciences in Biomedical and Communication Technologies*, 2011.

[4] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–18, Jan. 2012.

[5] I. Lee, O. Sokolsky, S. Chen, and J. Hatcliff, "Challenges and research directions in medical cyber physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 75–90, Jan. 2012.

[6] A. B. Sharma, F. Ivancic, A. Niculescu-Mizil, H. Chen, and G. Jiang, "Modeling and analytics for cyber-physical systems in the age of big data," *ACM Sigmetrics*, 2013.

[7] R. Mitchell and I. R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," *IEEE Trans. Depend. Secure Comput.*, vol. 12, no. 1, pp. 16–30, Jan.–Feb. 2015.

[8] R. Mitchell and I. R. Chen, "Adaptive intrusion detection of malicious unmanned air vehicles using behavior rule specifications," *IEEE Trans. Syst., Man, Cybern.*, vol. 44, no. 5, pp. 593–604, 2014.

[9] R. Mitchell and I. R. Chen, "Behavior rule based intrusion detection systems for safety critical smart grid applications," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1254–1263, 2013.

[10] R. Mitchell and I. R. Chen, "Effect of intrusion detection and response on reliability of cyber physical systems," *IEEE Trans. Rel.*, vol. 62, no. 1, pp. 199–210, Mar. 2013.

[11] R. Mitchell and I. R. Chen, "On survivability of mobile cyber physical systems with intrusion detection," *Wireless Personal Commun.*, vol. 68, pp. 1377–1391, 2013.

[12] I. R. Chen and D. C. Wang, "Analyzing dynamic voting using Petri nets," in *Proc. 15th IEEE Symp. Reliable Distributed Systems*, Niagara Falls, ON, Canada, Oct. 1996, pp. 44–53.

[13] I. R. Chen and D. C. Wang, "Analysis of replicated data with repair dependency," *Comput. J.*, vol. 39, no. 9, pp. 767–779, 1996.

[14] J. H. Cho, A. Swami, and I. R. Chen, "Modeling and analysis of trust management with trust chain optimization in mobile ad hoc networks," *J. Netw. Comput. Applicat.*, vol. 35, no. 3, pp. 1001–1012, 2012.

[15] U. S. DoE, Smart Grid, 2012 [Online]. Available: http://energy.gov/oe/services/technology-development/smart-grid

[16] J. H. Cho, I. R. Chen, and P. G. Feng, "Effect of intrusion detection on reliability of mission-oriented mobile group systems in mobile ad hoc networks," *IEEE Trans. Rel.*, vol. 59, no. 1, pp. 231–241, Mar. 2010.

[17] F. Bao, I. R. Chen, M. J. Chang, and J. H. Cho, "Trust-based intrusion detection in wireless sensor networks," in *Proc. IEEE Int. Conf. Communications*, Kyoto, Japan, Jun. 2011, pp. 1–6.

[18] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382–401, 1982.

[19] G. Ciardo, J. Muppala, and K. Trivedi, "SPNP: Stochastic Petri net package," in *Proc. 3rd Int. Workshop Petri Nets and Performance Models*, Washington, DC, USA, Dec. 1989, pp. 142–151.

[20] I. R. Chen and F. B. Bastani, "Effect of artificial-intelligence planning-procedures on system reliability," *IEEE Trans. Rel.*, vol. 40, no. 3, pp. 364–369, 1991.

[21] I. R. Chen, F. B. Bastani, and T. W. Tsao, "On the reliability of AI planning software in real-time applications," *IEEE Trans. Knowl. Data Eng.*, vol. 7, no. 1, pp. 4–13, 1995.

[22] D. Korzhyk, V. Conitzer, and R. Parr, "Solving stackelberg games with uncertain observability," in *Proc. 10th Int. Conf. Autonomous Agents and Multiagent Systems*, Taipei, Taiwan, May 2010.

Robert Mitchell received the B.S., M.S., and Ph.D. degrees in computer science from Virginia Polytechnic Institute and State University in 1997, 1998, and 2013, respectively.

Currently he is a programmer at Boeing. His research interests include security, mobile computing, sensor networks, embedded systems, and coding and information theory.

Ing-Ray Chen (M'89) received the B.S. degree from the National Taiwan University, Taipei, Taiwan; and the M.S. and Ph.D. degrees in computer science from the University of Houston.

He is a professor in the Department of Computer Science at Virginia Tech. His research interests include mobile computing, wireless networks, security, intrusion detection, trust management, real-time intelligent systems, and reliability and performance analysis.

Dr. Chen currently serves as an editor for IEEE Communications Letters, IEEE Transactions on Network and Service Management, *Wireless Communications and Mobile Computing, The Computer Journal, Wireless Personal Communications*, and *Security and Communication Networks*. He is a member of ACM.