Electricity Grid Cyber-Physical Security Risk Assessment Using Simulation of Attack Stages and Physical Impact

Yu-Cheng Chen, Vincent Mooney, and Santiago Grijalva
Georgia Institute of Technology
Atlanta, USA
ychen414@gatech.edu, mooney@ece.gatech.edu, sgrijalva@ece.gatech.edu

***Abstract*— Typical security risk assessment of cyber-physical systems measures the relative risk of individual components in the system with data from a domain expert. Such data may not come directly from the cyber-physical systems, but instead may come from the domain expert's knowledge. A challenging task is to assess the risk of a system-level cyber-attack (e.g., on an entire region with many substations), given the potential physical implications. This paper introduces a novel risk assessment tool that combines simulation of cyber-attack models and simulation of the potential physical impact in the power grid. The simulation involves probabilistic models of both the attack planning stage as well as the attack execution stage. The probabilities vary with estimated cyber-physical attacks and defensive postures. The proposed method provides insight into the risk of physical operational disruption caused by propagating cyber-attacks and provides strategies for their mitigation.**

I. Introduction

Cyber-physical systems such as the electricity grid can be subject to cyber-attacks that can result in significant societal disruption due to economic, physical, and even human loss. A recently released cybersecurity report by Siemens and the Ponemon Institute asserts that cyberthreats to utilities' operation technology are growing more serious. Of the 1,726 utility professionals surveyed, 54% expected at least one cyberattack on critical infrastructure within the next year [1, 2]. Two recent examples of cyberattacks on the power grid are the attacks on Ukraine's electricity grid in December 2015 and February 2016 [3]. The electricity industry has implemented and deployed standards, policies, and technologies to improve the cyber-security posture of the electricity grid. Electric utilities must assess the risk of cyber-attacks on the cyber-physical system (CPS) in order to develop better mechanisms to protect critical infrastructures.

Securing a CPS goes beyond securing the individual system components. A motivated adversary often uses the inter-dependency of vulnerabilities to carry out multistage attacks. While each phase of the attack may not by itself pose a serious threat to the corresponding component, the combined effect may be catastrophic. This paper introduces a novel risk assessment tool that combines simulation of cyber-attack models with the simulation of the potential physical impact in the power grid. The simulation involves probabilistic models of both the attack planning stage as well as the attack execution stage including potential physical impact.

II. Background and Prior Work

A. Power System Operational Security

Researchers have investigated the vulnerability of power grids by evaluating a set of plausible physical events, such as loss of transmission lines or generating units, and by evaluating their impact. Traditionally, power system operational security involves investigating *N*-1, *N*-2, and *N*-1-1 events [4]. In [5], a security assessment platform for the electric power system is presented which consists of risk data acquisition, risk identification, and analysis. However, the risk, generated as an output of the analysis, is not estimated based on any actual simulation of the response from the electric power system. Another similar work [6] uses the power system data such as the maximum voltage of each substation and the number of connection lines to calculate risk.

B. Power System Cyber-Security

A cyber-physical equivalent model for hierarchical control systems was proposed to evaluate the impact of inappropriate control commands on the power system [7]. A cyber-physical state estimation system was presented which identifies the maliciously modified set of measurements from power system sensors [8]. In [9], a mathematical framework for cyber-physical systems was built and used for attack detection and identification in power systems. Although these prior works are helpful, it is also important to assess the impact of the power grid when an attack is successful. In [10], a risk index evaluation system is proposed to cover the four cases of overload, cascading overload, low voltage, and voltage instability. In [11], a risk matrix technique is used to classify the risk of voltage collapse that would occur due to transmission line outage, but physical simulations are not carried out.

C. Attack Propagation

Prior work uses a Markov model to investigate attack propagation in the power grid [12]. This model defines a set of system states which model the success of the attacker in reaching various milestones of a kill chain, ultimately causing

a blackout. Corresponding probabilities are assigned to the state transitions. This Markov model approach can provide additional information to the system operator so that appropriate mitigation strategies against an attack can be implemented. In [13], probabilistic learning attacker, dynamic defender (PLADD) models the interaction of an attacker and a defender using game-theoretic analysis.

In this paper, we use our recently published hybrid attack model [14] to model attack scenarios. The hybrid attack model is a combination of the Markov model and the PLADD model. First, we split an attack into two stages: preparation and execution. The preparation stage describes attacker actions that involve gathering information. The execution stage describes attacker actions that involve using the gathered information to attack the power grid. The hybrid attack model uses PLADD games to model the information required to initiate an attack. The hybrid attack model uses the Markov states to model actions representing the attacker's process to execute an attack on the power grid.

III. Attack Scenarios

In order to illustrate the methodology, let us consider a four-bus power system where each bus is located at a substation. Substation 1 and Substation 4 are generating power while Substation 2 and Substation 3 have loads that consume power. We assume that each substation has one remote terminal unit (RTU) that collects data from the substation sensors and can execute control center commands. Specifically, RTUs are capable of opening/closing breakers on the transmission lines.

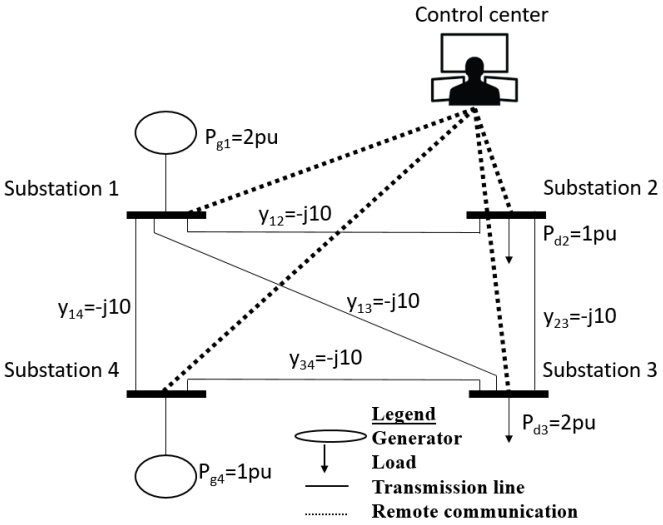

Fig. 1. Four-bus power grid system.

The attacker's attack plan for a single substation is shown in Fig. 2. We assume that the attacker first gathers all necessary information prior to executing an attack on the substation. We assume the attacker requires access to a "Vulnerability Report," "RTU credentials," and "IP address of the RTU." The Vulnerability Report represents a set of relevant information about the system that is useful to implement the attack. The Vulnerability Report consists of other information regarding items such as (1) equipment type and/or design, (2) criticality of load, and (3) number and/or types of customers. We assume the Vulnerability Report for the entire power grid is stored on a utility engineer's computer. We further assume that the attacker will need to run password cracking software to gain access to the vulnerability report. The attacker would use a brute force attack on each RTU to gain control. Each RTU is assumed to have different login credentials. We also assume the attacker does not want the control center to remotely control a substation that is under the attacker's control. Therefore, the attacker would execute Address Resolution Protocol (ARP) cache poisoning [15] to prevent communication between the control center and the substation. After the attacker has gathered the necessary information, the attacker executes the attack by the following steps: (1) breaching the substation's room locked door, (2) accessing the RTU, (3) disabling communication between the substation and control center and (4) opening breaker(s) of transmission lines at a substation. We also assume multiple substations can be attacked simultaneously, given that the attacker has all the necessary information. In our attack scenario, we assume the attacker's goal is to maximize the loss of power to ccustomers in the grid. The attacker can attack any of the four substations shown in Fig. 1. In addition, we also consider the scenario where the attacker knows that Substation 1 and Substation 4 generates power in the power grid, and the attacker decides to cause a blackout for the entire grid.

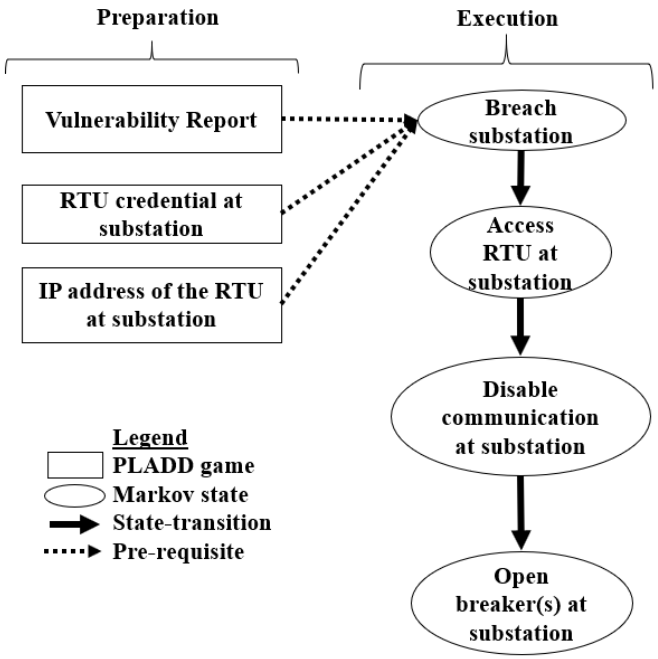

Fig.2. Attacker's attack plan for a single substation

IV. Integration of Hybrid Attack Model and Grid Impact Analysis

Unlike the prior risk assessment work such as *N*-1 discussed in Section II, our risk assessment of the power grid will be specific to the attacker's goal. An example attacker's goal could be to overload the transmission line with the intent of activating power relays that would disconnect the line and cause loss of load. Our risk assessment tool is be able to analyze and rank the risk of each substation in Fig. 1. The security analysts will need to determine the type of attack that will cause the most damage,

and the tool will rank the risk of each substation based on the specific attack. For example, the security analyst in a populated region may determine that a blackout is the most devastating attack that can be done by the attacker because the number of customers is high. On the other hand, in a secluded region, the security analysts may determine that overloading a transmission line is more costly than a small neighborhood blackout. In order to create a risk assessment tool that incorporates a realistic grid response to an attack, we have integrated power flow analysis with the hybrid attack model.

We use a simplified DC power flow model as shown in Equation 1.

$$\theta = [B']^{-1} P_{injection} \quad (1)$$

where B' is the reduced susceptance matrix describing the topology and parameters of the power grid. θ is the bus voltage angle with respect to the system reference (slack) bus. $P_{injection}$ is the vector of bus power injections at a given point in time. A positive entry in the $P_{injection}$ vector represents power produced by a generator, and a negative value corresponds to the power consumed by a load at that bus. The flow of active power on a given transmission line between buses j and k is

$$P_{jk} = B_{jk}(\theta_j - \theta_k) \quad (2)$$

where B_{jk} is the value in susceptance matrix B' located at row k and column j. θ_k and θ_j are the angle of active power at bus k and bus j respectively.

We note that depending on the topology of the system, the disconnection of one or more lines can cause loss of load at more than one substation, e.g. substations have a radial topology from generation to loads, where the loads are downstream with respect to the attacked substation.

As described in [14], the hybrid attack model combines the advantages of the PLADD model's timing information to improve the Markov model's ability to assess the security risk against a specific attack. In this paper, we evaluate the risk of each substation against a specific attack. Similar to the prior work [10] [11] on risk assessment described in Section II.B, we define risk as the probability of successful attack multiplied by the severity of the damage. The probability of a successful attack (P) is calculated as shown in Equation 3.

$$P = \frac{Number\ of\ days\ the\ attacker\ has\ the\ ability\ to\ open\ breakers}{Number\ of\ days\ in\ the\ simulation} \quad (3)$$

It is noteworthy to point out that prior work on risk assessment requires a domain expert to input the severity. In this paper, we use the percentage of power loss to represent the severity of the damage. The severity is calculated as shown in Equation 4.

$$Severity = \frac{Number\ of\ loads\ not\ satisfied}{Total\ load\ in\ the\ grid} \quad (4)$$

The severity of each test case is shown in Table I. The equation to calculate risk is shown in Equation 5.

$$Risk = P * Severity \quad (5)$$

TABLE I. THE SEVERITY OF DAMAGE FOR EACH TEST CASE

Test cases	Severity (percentage of power loss)
1	0.66
2	0.33
3	1
4	0.33
5	1

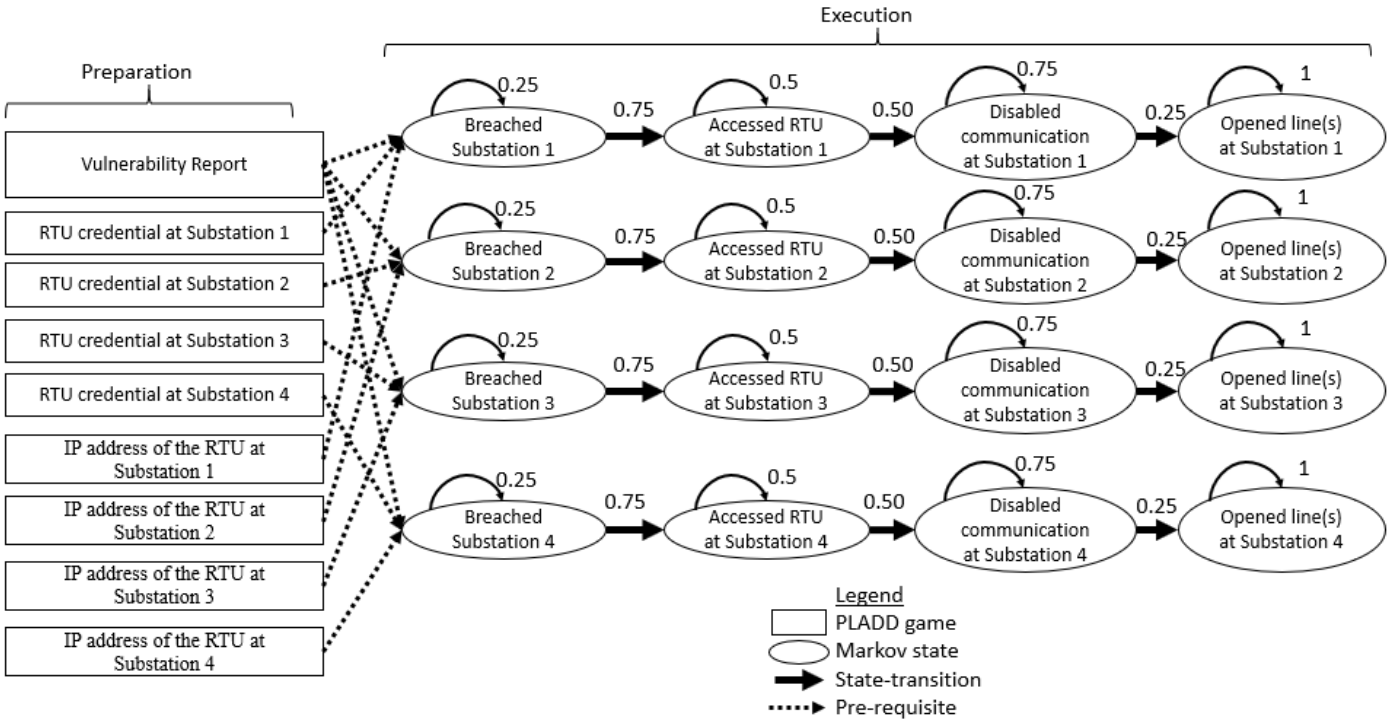

Fig. 3. Hybrid attack model of the four-bus system

V. Experiment Results

As described in [14], each PLADD game is modeled by two parameters τ and $f_{base}(t)$, where τ is the defender's period to periodically execute a take move and $f_{base}(t)$ is the probability density function describing the attacker's time-to-success of an attack. In this paper, we assumed $f_{base}(t)$ is an exponential distribution, so $f_{base}(t) = \frac{1}{\mu} e^{-\frac{t}{\mu}}$ where μ is the average. Therefore, instead of using τ and $f_{base}(t)$ to model a PLADD game, we will be using τ and μ to model a PLADD game. Fig. 3 shows the hybrid attack model of the four-bus system. The probability of transitioning from one state to the next is shown on the state transition edges in Fig. 3. Once the attacker has gathered all the necessary information in the preparation stage, the attacker will move forward by attacking the power grid. In the execution stage, we assume the attacker will (1) breach the substation fences, (2) access the RTU at the substation, (3) disable communication at the substation, and (4) open line(s) at the substation. For simplicity, we assume the attacker is making six state transitions in Fig. 3 per day. Our risk assessment tool can change the number of transitions per day by changing a single input parameter. The PLADD parameters used to model the four-bus system are shown in Table II. We assume the following items with regard to Table II.

- The vulnerability report is changed every 180 days (approximately six months). The average time-to-success of the attacker's password cracking software is 90 days (approximately three months).
- The RTU credentials at Substation 1, Substation 2, and Substation 3 are changed every 90 days. For Substation 4, we assume the RTU is a newer model as compared to other substations, which allows the operator to change the login credentials every 45 days. The average time-to-success of the attacker's brute force attack on the RTUs in Substation 1, Substation 2, Substation 3, and Substation 4 is 45 days.
- The IP addresses of the RTUs at Substation 1, Substation 2, and Substation 3 are changed every 360 days. For Substation 4, we assume the RTU is a newer model as compared to other substations, which allows the operator to change the login credentials every 180 days. We assume the attacker needs to analyze network packets to figure out the IP addresses of RTUs. Therefore, the average time required to figure out the IP address of an RTU from analyzing network packets is 180 days.

Our experiment is implemented in Matlab. In our experiment, we simulate an attack using the hybrid attack model and calculate the probability of success using Equation 3. When an attack is successful against a substation, the attacker disconnects the transmission line(s) connected to the substation. Next, we use the DC power flow analysis to calculate the power on each transmission line and at each substation after a successful attack from the attacker. As a result, we calculate the loss of power due to a successful attack. Next, we use the DC power flow analysis to calculate the loss of power as a result of a successful attack from the attacker. We have implemented five test cases for comparison of risk. The DC power flow parameters are shown in Table III.

- Test case 0: Normal power grid operation (base case).
- Test case 1: Attacker attempts to disconnect Substation 1 from the grid, which forces the operator to shed 2 pu load between Substation 2 and Substation 3.
- Test case 2: Attacker attempts to disconnect $line_{12}$ and $line_{23}$ in Substation 2, which causes a power loss of 1 pu.
- Test case 3: Attacker attempts to disconnect $line_{13}$, $line_{23}$, and $line_{34}$ in Substation 3, which causes a power loss of 2 pu.
- Test case 4: Attacker attempts to disconnect $line_{14}$ and $line_{34}$ in Substation 4, which forces the operator to shed 1 pu load between Substation 2 and Substation 3.
- Test case 5: Attacker attacks Substation 1 and Substation 4 to cause blackouts at Substation 2 and Substation 3.

TABLE II. PLADD PARAMETERS USED TO MODEL THE FOUR-BUS SYSTEM

PLADD game type	τ (day)	μ (day)
Vulnerability report	180	90
RTU credentials at Substation 1, Substation 2, Substation 3	90	45
IP addresses of the RTU at Substation 1, Substation 2, Substation 3	360	180
RTU credential at Substation 4	45	45
IP address of the RTU at Substation 4	180	180

TABLE III. DC POWER FLOW PARAMETERS

Test case	y_{12}	y_{13}	y_{14}	y_{23}	y_{34}	P_1	P_2	P_3	P_4
0	10	10	10	10	10	2	-1	-2	1
1	0	0	0	10	10	2	-1	-2	1
2	0	10	10	0	10	2	-1	-2	1
3	10	10	10	10	0	2	-1	-2	1
4	10	10	0	10	0	2	-1	-2	1
5	0	0	10	10	0	2	-1	-2	1

The risk calculation of test cases 1-5 are shown in Table IV. The simulation runs for 720 days (or approximately two years) with the result that Substation 3 has the highest risk and Substation 4 has the lowest risk. Fig. 4 through Fig. 6 show the simulation results of the hybrid attack model. Fig. 4 shows simulation results of the hybrid attack model for attacking Substation 3. Fig. 5 shows simulation results of the hybrid attack model for attacking Substation 4. Fig. 6 shows simulation results of the hybrid attack model for attacking Substation 1 and Substation 4. Each PLADD game is represented as a state where "0" means the defender has control of the node and "1" means the attacker has control of the node. The top three plots in each of Fig. 4 through Fig. 6 show the state of each PLADD game with respect to time, and the bottom subplot in each shows the product of all PLADD game states with respect to time. The product of all PLADD game states is calculated by multiplying the state of each PLADD game. When the product of all PLADD game states is equal to 1, the

attacker has all the necessary information to execute an attack on the grid. By comparing the product of all PLADD states in Fig. 5 and Fig. 6, we can see that the attacker is spending less time in the execution stage when the attacker is attacking Substation 4. This is an expected result because we assumed the RTU at Substation 4 is a newer model as compared to the RTUs at Substation 1, Substation 2, and Substation 3. The operator can reset the RTU's password and IP address at Substation 4 at a faster pace. By comparing the product of all PLADD states in Fig. 4 and Fig. 5 against Fig. 6, we can see that it is more difficult to attack two substations instead of one substation.

TABLE IV. RISK CALCULATION OF TEST CASE 1-5

Test case	Number of successful attacks in simulation	Total number of days in simulation	Probability of successful attack	Severity (Percentage of power loss)	Risk
1	164	720	0.2278	0.66	0.1503
2	217	720	0.3014	0.33	0.0995
3	272	720	0.3778	1.00	0.3778
4	49	720	0.0681	0.33	0.0225
5	40	720	0.0556	1.00	0.0556

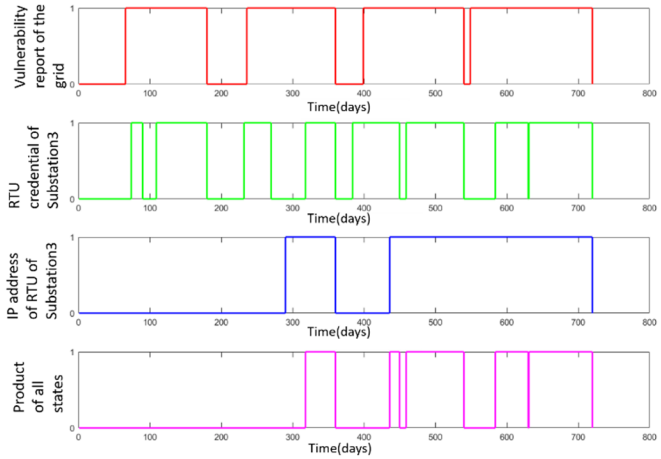

Fig. 4. A hybrid attack model simulating an attack on Substation 3.

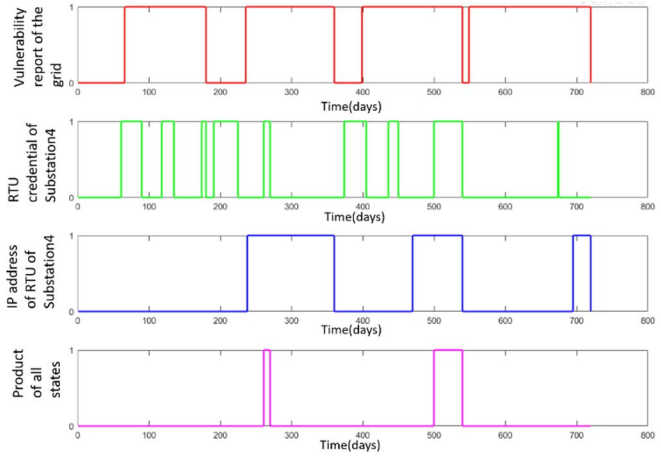

Fig. 5. A hybrid attack model simulating an attack on Substation 4.

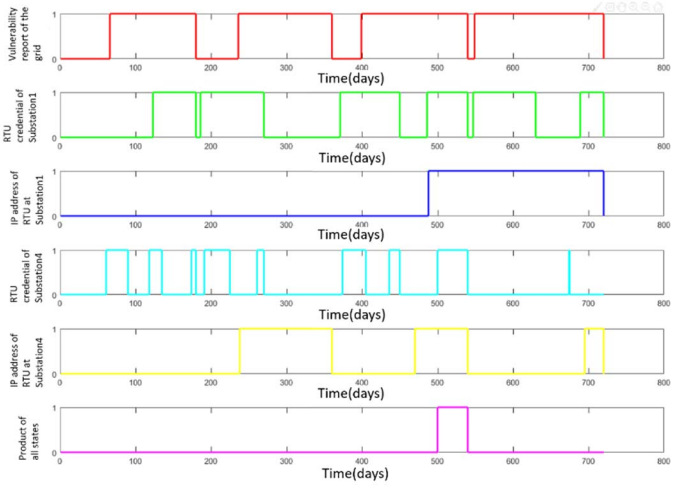

Fig. 6. A hybrid attack model simulating an attack on Substation 1 and Substation 4 simultaneously.

VI. DISCUSSION

The risk calculation in Table IV provides a way to compare (1) the probability of success of different attacks, (2) risk of individual substations, and (3) the risk of a combination of substations. Intuitively, since the parameters used to model Substation 1, Substation 2, and Substation 3 are the same, the probability of success for attacking Substation 1, Substation 2 and Substation 3 should be the same. However, as described in [14], the attacker's time-to-successful attack is based on an exponential distribution function. The exponential distribution function is used as an input to a pseudo-random sampling technique called inverse transform sampling [16]. The inverse transform sampling technique generates pseudo-random numbers following an exponential distribution. The random number generation is the reason that the probability of a successful attack on Substation 1, Substation 2, and Substation 3 are different, even though the PLADD parameters used to model Substation 1, Substation 2, and Substation 3 are the same. By comparing test case 1, test case 2, and test case 3, we can see that the risk of test case 3 is the highest, and the risk of test case 2 is the lowest, which reflects the severity of each attack. By comparing test case 1 (attack on Substation 1) and test case 4 (attack on Substation 4) with test case 5 (attack on both Substation 1 and Substation 4), we can see that the probability of success for simultaneously attacking Substation 1 and Substation 4 is less than the probability of success for attacking Substation 1 and Substaion4 independently. Therefore, even though the severity of test case 5 is higher than test case 1 and test case 4, the risk of test case 5 is less than test case 1 and test case 4. In addition, due to the assumption that Substation 4 has a newer RTU model, the probability of a successful attack on Substation 4 is lower than the other substations. Of course, in a practical application of the approach proposed in this paper, many additional test cases beyond the five explained here would need to be carried

out to provide proper coverage of the attack surface, and such coverage invariably relies on expert domain knowledge. With the information in Table IV, security analysts can relate a risk value to the percentage of power loss after a successful attack to a substation. For example, the damage associated with test case 1's risk is load shedding of 2 pu between Substation 2 and Substation 3. In this paper, a risk value is related to parameters associated with the defender, the attacker, and the power loss of the grid. For example, security analysts may decide that the highest acceptable risk is 0.1503 (test case 1's risk). This means the security analysts need to reduce the risk of Substation 3 by (1) increasing the frequency of credential resets for Substation 3, (2) increasing the average time the attacker needs to perform a successful attack (e.g., increase length of passwords) and/or (3) implementing redundant transmission lines to reduce the severity of Substation 3.

VII. Conclusion

Our simulations show that it is feasible to measure the risk of a power grid subject to a specific cyber-physical attack. A potential application of our risk assessment tool is in the field of cyber-physical attack-resilience in DC Microgrid. We use power flow analysis to calculate the loss of load to customers after an attack is successful. Security analysts could use the proposed method to map the risk of substations in the power grid to the actual loss of load due to an attack. Security analysts will also be able to determine which substation in the power grid is at the most risk from a specific attack. This information is important in order to prioritize mitigation actions. The security analysts will also be able to provide information regarding the severity and impact of the risk. If the risk is not acceptable, then the security analysts can assist in finding location(s) in the power grid to which to devote more resources to improve security.

References

[1] L. Simonovich, "Are utilities doing enough to protect themselves from cyberattack?," World Economic Forum, 5 January 2020. [Online]. Available: https://www.weforum.org/agenda/2020/01/are-utilities-doing-enough-to-protect-themselves-from-cyberattack/. [Accessed 20 May 2020].

[2] L. Simonovich, "Caught in the Crosshairs: Are Utilities Keeping Up with the Industrial Cyber Threat?," Siemens, Houston, Texas, 2020.

[3] R. M. Lee, M. J. Assante and T. Conway, "Analysis of the cyber attack on the Ukraine power grid," SANS Ind. Control Syst., 2016.

[4] Z. Qiqi, S. Gang, F. Yuyao and S. Yun, "Network topology of urban grid considering N-1-1 criterion," *12th IET International Conference on AC and DC Power Transmission (ACDC 2016)*, Beijing, 2016.

[5] Y. Yu and W. Lin, "Study on the security assessment paltform for electric power secondary system," *International conference on Power System technology*, Chongqing, 2006.

[6] F. Farzan, M. A. Jafari, D. Wei and Y. Lu, "Cyber-related risk assessment and critical asset identification in power grids," *ISGT*, Washington, DC, 2014.

[7] S. Xin, Q. Guo, H. Sun, B. Zhang, J. Wang and C. Chen, "Cyber-phyiscal modeling and cyber-contingency assessment of hierarchical control systems," *IEEE Trans. on Smart Grid,* vol. 6, no. 5, pp. 2375-2385, 2015.

[8] S. Zonouz, K. M. Rogers, R. Berthier, R. B. Bobba, W. H. Sanders and T. J. Overbye, "SCPSE: Security-oriented cyber-physical state estimation for power grid critical infrastructures," *IEEE Trans. on Smart Grid,* vol. 3, no. 4, pp. 1790-1799, 2012.

[9] F. Pasqualetti, F. Dorfler and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Auto. Control, ,* vol. 58, no. 11, pp. 2715-2729, 2013.

[10] M. Ni, J. D. McCalley, V. Vittal and T. Tayyib, "Online risk-based security assessment," *IEEE Transactions on Power Systems,* vol. 18, no. 1, pp. 258-265, 2003.

[11] N. Aminudin, N. M. Ramli, M. Marsadek, N. M. Ramli and T. K. A. Rahman, "Classification of risk of voltage collapse using risk matrix," *2016 IEEE International Conference on Power System Technology (POWERCON)*, Wollongong, NSW, 2016.

[12] V. Chukwuka, Y. Chen, S. Grijalva and V. Mooney, "Bad Data Injection Attack Propagation in Cyber-Physical Power Delivery System," in *Clemson University Power Systems Conference (PSC)*, Charleston, SC, USA, 2018.

[13] S. Jones, A. Outkin, J. Gearhart, J. Hobbs, J. Siirola, C. Philips, S. Verzi, D. Tauritz, S. Mulder and A. Naugl, "Evaluating Moving Target Defense with PLADD," Sandia National Laboratories, 2015.

[14] Y. Chen, T. Gieseking, D. Campbell, V. Mooney and S. Grijalva, "A Hybrid Attack model for Cyber-Physical Security Assessment in Electricity Grid," *IEEE Texas Power and Energy Conference (TPEC)*, College Station, Texas, USA, 2019.

[15] M. Al-Hemairy, S. Amin and Z. Trabelsi, "Towards more sophisticated ARP Spoofing detection/prevention systems in LAN networks," *International Conference on the Current Trends in Information Technology (CTIT)*, Dubai, 2009.

[16] M. Bonakdarpour, "Inverse Transform Sampling," 02 February 2016. [Online]. Available: https://stephens999.github.io/fiveMinuteStats/inverse_transform_sampling.html. [Accessed 20 May 2020].