CyberDep: Towards the Analysis of Cyber-Physical Power System Interdependencies Using Bayesian Networks and Temporal Data

Leen Al Homoud*§, *Student Member, IEEE*, Katherine Davis*, *Senior Member, IEEE*,
Shamina Hossain-McKenzie‡, *Member, IEEE*, Nicholas Jacobs‡, *Member, IEEE*,
*Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA
‡Sandia National Laboratories, Albuquerque, NM, USA
§leen.alhomoud@ieee.org

***Abstract*—Modern-day power systems have become increasingly cyber-physical due to the ongoing developments to the grid that include the rise of distributed energy generation and the increase of the deployment of many cyber devices for monitoring and control, such as the Supervisory Control and Data Acquisition (SCADA) system. Such capabilities have made the power system more vulnerable to cyber-attacks that can harm the physical components of the system. As such, it is of utmost importance to study both the physical and cyber components together, focusing on characterizing and quantifying the interdependency between these components. This paper focuses on developing an algorithm, named *CyberDep*, for Bayesian network generation through conditional probability calculations of cyber traffic flows between system nodes. Additionally, *CyberDep* is implemented on the temporal data of the cyber-physical emulation of the WSCC 9-bus power system. The results of this work provide a visual representation of the probabilistic relationships within the cyber and physical components of the system, aiding in cyber-physical interdependency quantification.**

***Index Terms*—Cyber-Physical Interdependencies, Cyber-Physical Power Systems, Graph Theory, Bayesian Networks, Dependency Graphs, Temporal Data**

I. Introduction

Over the past decade, power systems have become increasingly recognized as cyber-physical systems. The importance of understanding the interdependency between cyber and physical components is highlighted by the many new developments in the grid, such as, but not limited to, renewable energy integration, distributed energy generation monitoring and control, and new cyber-security technology. Specifically, a power grid's Supervisory Control and Data Acquisition (SCADA) system allows for the monitoring and control of physical components in the system and communicates with devices through a variety of protocols, such as the Distributed Network Protocol 3 (DNP3) [1]. This protocol is one of the most widely used in electric power utilities. Such developments have made the power grid more vulnerable to cyber-attacks that target the physical components of the system at the generation, transmission, and distribution levels. Two infamous threats are the Industroyer and Industroyer2 malware that affected Ukraine in 2016 and 2022, respectively [2]. Both of these threats targeted electrical substations in the country, with the Industroyer malware sending SCADA commands to the field devices resulting in an hour-long power outage across Ukraine.

As such, it is now of utmost importance to analyze and study the power system as both a cyber and physical system, while taking into account how the cyber and physical components of the system are interdependent. There is a lot of emerging literature focused on understanding cyber-physical power system interdependencies [3]–[6]. In [3], the authors study the interdependency relationship between the physical power grid and its corresponding communication network when dealing with and mitigating cascading failures. Through numerical simulations of a cyber-attack on a cyber-physical power system model, the authors found that power systems divide into clusters when facing cascading failures. These results showed that there is a correlation between system robustness and cluster size, proving that these cyber-physical clusters are still interdependent of each other, but operating separately. In [4], the authors claim that "interdependence is an intrinsic feature of cyber-physical systems." The authors back up this claim by characterizing cyber-physical interdependencies using correlation metrics aimed at predicting the propagation of failure following a cyber-attack on the network. Huang et al. [5] also study interdependencies concerning cascading failures following a mathematical estimation approach using concepts of graph theory. Other applications that utilize cyber-physical interdependency analysis in power grids include improving power system reliability modeling [6] and developing cyber-physical resiliency metrics [7], [8].

In addition to the literature detailed above, Bayesian Networks have also been used in cyber-physical power systems for many applications including, but not limited to, attack graph generation [9], cyber threat mitigation [10], scalable anomaly detection [11], and risk analysis and assessment [12], [13]. Sahu et al. [9] focused on developing a Bayesian attack graph and updating it through the use of constraint-based structural learning methods that focus on scalability and

This work was supported by the Sandia Laboratory Directed Research and Development Project #229324 and the US Department of Energy under award DE-CR0000018.

979-8-3503-7240-3/24/$31.00 ©2024 IEEE

accuracy. In [10], the authors develop a quantitative framework using Bayesian networks to define all possible vulnerabilities and optimize this framework to achieve mitigation of cyber-physical attacks, while Krishnamurthy et al. [11] focused on creating Bayesian networks of power systems to study the different cyber-physical relations between the nodes to achieve anomaly detection focused on power system scalability.

It is also important to note that this work is part of a larger effort aimed at characterizing and quantifying cyber-physical power system interdependencies [14]–[16]. Much of the ongoing work is focused on the development and use of a variety of graph clustering methods that aid in characterizing cyber and physical disturbances and cyber-physical interdependencies. Therefore, the work in this paper focuses on continuing these efforts through the development of a Bayesian network generation algorithm that inputs temporal data generated from the earlier work in [17] and outputs graph visualizations of the probabilistic relationship between different nodes in a cyber-physical power system model.

With that being said, the contributions of this paper are as follows:

1) Development of a Bayesian network generation algorithm through the use of temporal data and conditional probability calculations of cyber traffic flows between system nodes.
2) Application of this algorithm on the temporal data of the cyber-physical emulation of the WSCC 9-bus system [18] under physical, cyber, and cyber-physical disturbances.
3) Visualization of the probabilistic relationships between the different system nodes, aiding in cyber-physical interdependency quantification.

II. Methodology

In this section, we first describe the temporal dataset generated by the earlier work in [17] and list the physical, cyber, and cyber-physical threat vectors that make up the different disturbance scenarios. Then, we focus on conceptualizing Bayesian networks and detailing the development of the generation of such networks, specifically known as Dependency Graphs.

A. Disturbances and Dataset Description

The HARMONIE [17] project focused on developing a cyber-physical response engine that generates real-time cyber-physical power system mitigations through a machine learning classification framework and automated remedial action schemes (RAS). The techniques developed were tested in a cyber-physical emulation environment built using a real-time digital simulator (RTDS) and SCEPTRE™ [19]. SCEPTRE™ [19] is a modeling and emulation platform developed by Sandia for emulating Industrial Control Systems (ICS). It allows for the modeling and emulation of different virtual and hardware devices, such as, but not limited to, switches, servers, and relays. It also supports power system simulations and ICS communication protocols such as DNP3.

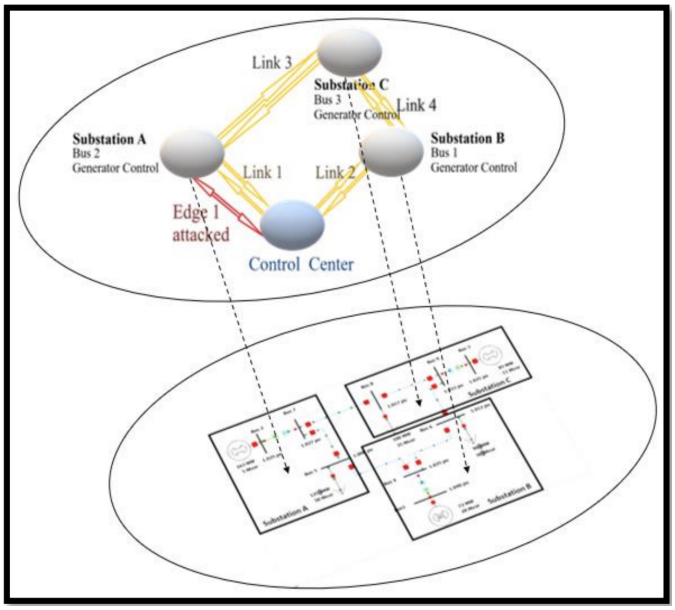


Fig. 1: Cyber-physical mapping of the WSCC 9-bus system [17].

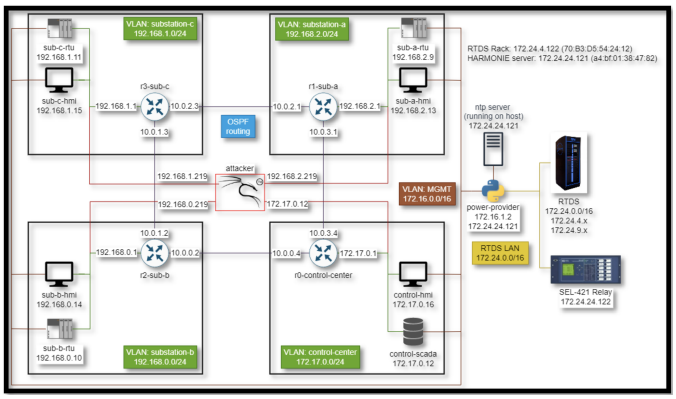


Fig. 2: Network diagram of the WSCC 9-bus emulation environment [17].

The data used in this paper was generated as part of the different experiments that were run in the emulation environment on the WSCC 9-bus power system [18], where the cyber-physical mapping of the 9-bus system is shown in Figure 1 and the network diagram is shown in Figure 2. In Figure 1, it can be observed that the WSCC 9-bus system is divided into three substations, with Substation A containing Bus 2, Substation B containing Bus 1 (slack bus), and Substation C containing Bus 3. The fourth substation in the cyber-physical emulation is the control center, which contains the SCADA system that sends commands to the field devices. In the environment, the WSCC system is emulated as a 4-substation network, with a router connecting each substation to the rest of the network, as shown in Figure 2.

The three disturbances that were tested in this environment are [17]:

1) Cyber: The cyber disturbance consisted of a Denial-of-Service (DOS) intrusion. The mitigation implemented for this threat included using firewall rules to block adversary communication.
2) Physical: The physical disturbance included the loss of a generator and a branch that led to line overloading. The mitigation implemented is load shedding at two different buses using an automated remedial action schemes (AutoRAS) algorithm [17], [20], [21].
3) Cyber-Physical: This disturbance is a combination of the above disturbances with both mitigation strategies implemented.

The dataset contains four cyber-physical disruption scenarios based on the three disturbances listed above. These scenarios are run three times each and are as follows:

1) Baseline: Normal system operations.
2) DOS: This scenario includes only the cyber disturbance with no physical disruptions.
3) No Mitigation: This scenario includes a physical disturbance as well as a cyber one that affects load shedding.
4) Mitigation: This is the same as the No Mitigation scenario with an addition of the firewall rules put in place to block the cyber-attack.

B. Dependency Graph Generation

Dependency graphs (DGs) are a type of Bayesian network that helps represent the different cyber and physical system characteristics during normal operating conditions and under threats. Dependency graph (DG) conceptualization is provided in [22], where the authors focused on developing a cyber-physical resiliency metric using graph theory concepts. For this paper, we will focus on DGs to help quantify cyber-physical interdependencies. DGs are generated through the conditional probability calculations of the frequency of communication between the different nodes using the following formula [22]:

$$P(x|P(x)) = 1 - \prod_{p_x^i \in P(x)} \left(1 - \mathbf{1}_{(p_x^i)} \times P(p_x^i \rightarrow x)\right)$$

where,

$P(x|P(x))$: Probability of *x* given *P(x)*

$\mathbf{1}_{(p_x^i)}$: Indicator function, which is 1 if the condition in parentheses holds and 0 otherwise

$P(p_x^i \rightarrow x)$: Probability of information flow from p_x^i to *x*

A DG captures the relationships between the different files and processes in a network, which depend on whether there is data flow between two nodes. As such, a DG implies that if there is traffic moving from object o_i to o_j, then object o_j is dependent on object o_i. This dependency is represented by an edge on the graph, $o_i \rightarrow o_j$. In this example, the dependency relationship is characterized by three components, which are the source, o_i, the sink, o_j, and the security contexts, cyber traffic information between nodes o_i and o_j. The nodes of a DG are modeled as binary random variables, and the edges

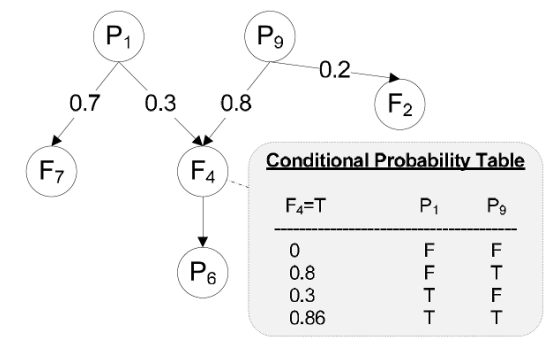


Fig. 3: Sample dependency graph obtained from [22].

are labeled with the frequency of communication between two different nodes, which is the calculated probability dependent on the number of system calls between each of the nodes.

System calls, $syscalls$ in short, are the communication requests and responses made between each node. For the DNP3 communication protocol, the $syscalls$ under consideration are Request Link Status, Read, Respond, and Direct Operate commands, explained in more detail in the DNP3 protocol primer [1]. A sample of a dependency graph can be seen in Figure 3 [22]. The conditional probability that File F4 would be affected if a cyber-attack were to affect either Process P1 or P9 is given by [22]:

$$P(F4|P1, P9) = 1 - (1 - 0.3) \times (1 - 0.8) = 0.86$$

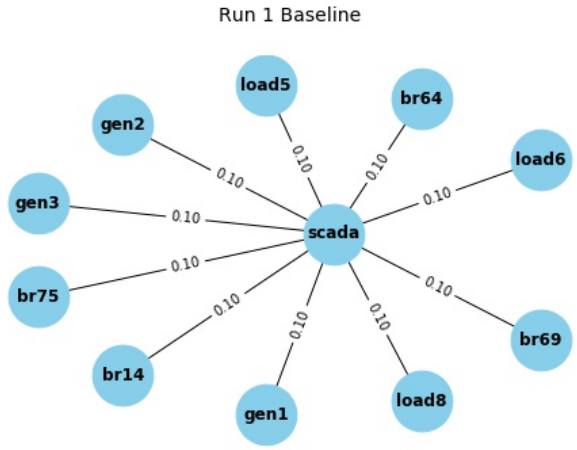
Similarly, the probability that File F2 would be affected if Process P9 is affected is 0.2, and the probability that File F7 would be affected if Process P1 is affected is 0.7.

Algorithm 1 shows the steps for generating the DGs. The datasets collected from the emulation are in JSON file format. The first step is to load the files, and then filter the traffic out for DNP3 data. DNP3 traffic is selected because it collects information on the physical components of the network as well as cyber information, thus providing a better insight into cyber-physical interdependencies. Once the input data is processed, the IP addresses are then mapped to the device names using the network topology information. The frequency of communication is then counted, and the conditional probability is then calculated. Four graphs for each of the runs are generated, as a result, for each of the four scenarios.

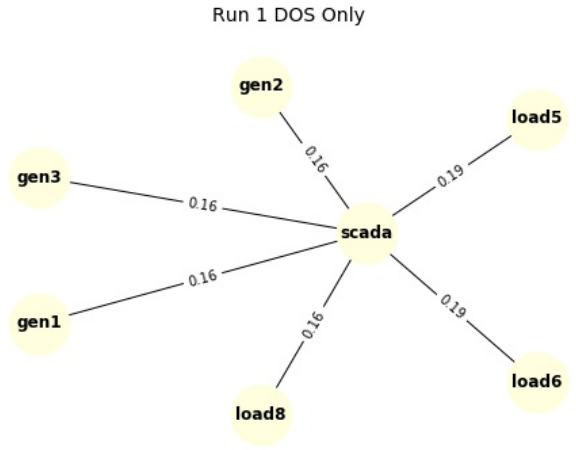
III. Results and Discussions

In this section, we will discuss the results for each of the three experimental runs and their respective four cyber-physical disruption scenarios. Figures 4, 5, and 6 display the results for runs 1, 2, and 3, respectively.

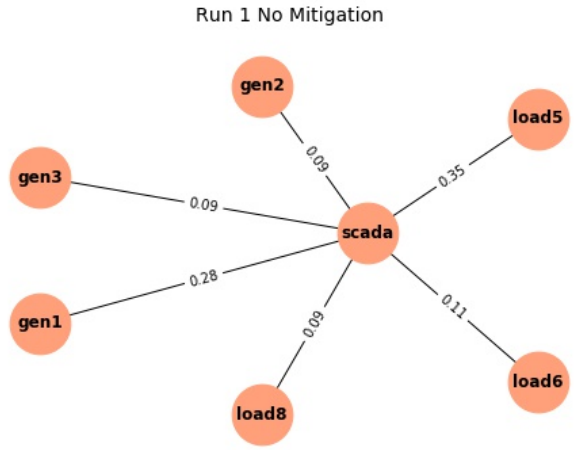
Across all three experimental runs, the baseline graphs exhibited similar patterns of behavior in which the probabilities of all the edges were equal. The probabilities amounted to 0.1 for each edge in run 1 and run 2 and 0.17 for each edge in run 3. A crucial observation is that the DOS Only, No Mitigation, and With Mitigation scenarios behaved similarly across runs



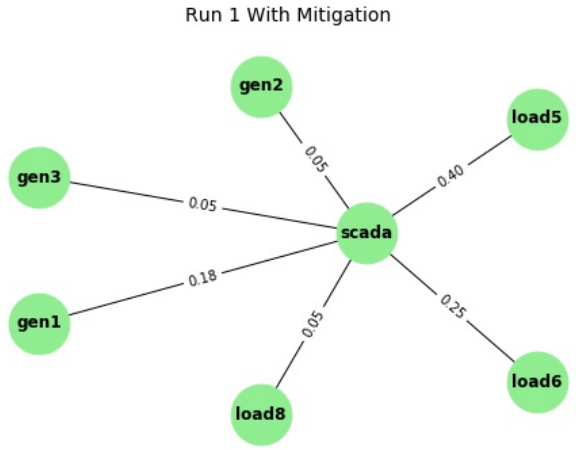
(a) Baseline results for Run 1.



(b) DOS results for Run 1.



(c) No mitigation results for Run 1.



(d) Mitigation results for Run 1.

Fig. 4: Results for all four scenarios for the first run of the experiments in the dataset. The edges are labeled with the probabilities calculated using the temporal data.

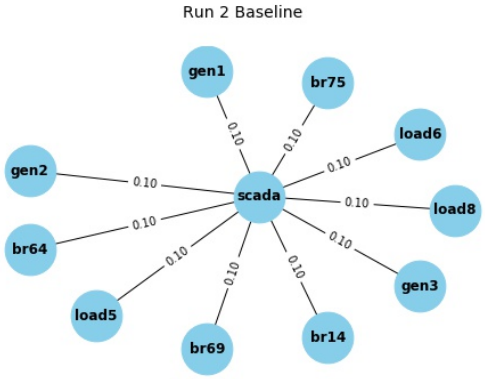
Algorithm 1: Dependency Graph Generation

Data: Raw temporal data collected in [17].
Result: Four graphs per experiment run that visualize the probabilistic relationship between the different nodes in the system.

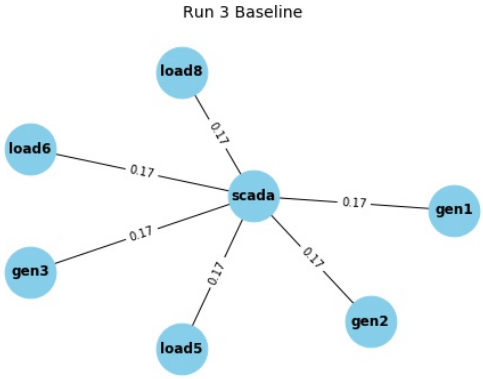
1 Load JSON files of temporal data.
2 Filter for DNP3 traffic data.
3 Convert JSON files to CSV format.
4 Map IP addresses with device names using the network topology.
5 Count the number of times each component communicated with the control center's SCADA system.
6 Calculate the conditional probability to quantify the frequency of communication between the nodes.
7 Plot the graphs for each of the four scenarios for all three runs and label the edges with the probabilities.

1 and 2; however, the DOS Only scenario in run 3 behaved differently. For runs 1 and 2, the DOS Only scenario shows that the highest probabilities are for the edges connecting each of loads 5 and 6 to the SCADA node. This result makes sense as this scenario consists of a DOS threat through DNP3 increasing the amount of packets traveling between the objects affected. It is also important to note that while this intrusion was cyber in nature, we were able to see the relation and the effect of a purely cyber-attack on two physical components in the network, loads 5 and 6. For the DOS Only scenario for run 3, the opposite is observed. The edges connecting loads 5 and 6 to the SCADA node have the lowest probabilities. This could be due to the fact that the DOS threat in this scenario was not implemented for the same duration of time as the other two runs. As such, further analysis of the network topology and experimental setup would be required to interpret this result.

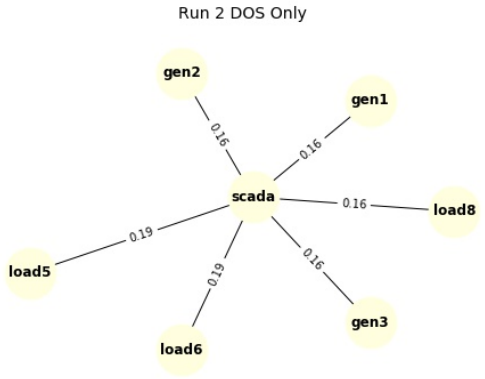
Moving onto the No Mitigation and With Mitigation scenarios for all three runs, similar patterns and probabilities were observed for the edges in the graphs. Specifically, an important



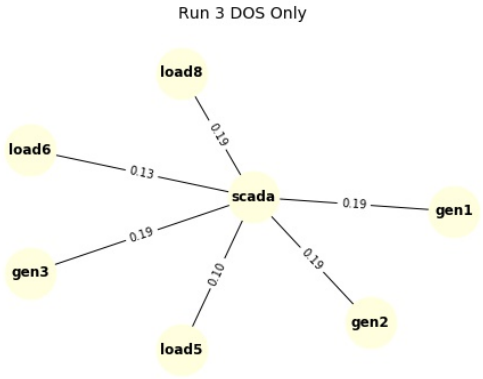
(a) Baseline results for Run 2.



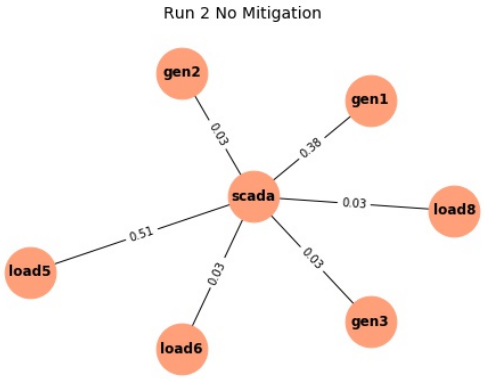
(a) Baseline results for Run 3.



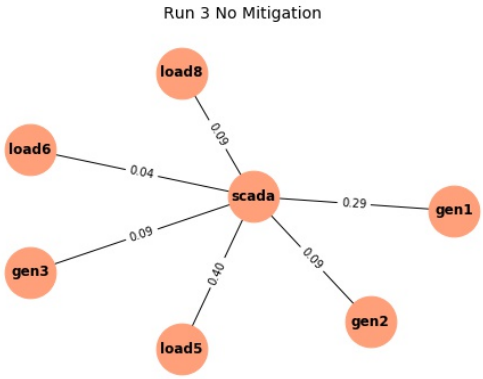
(b) DOS results for Run 2.



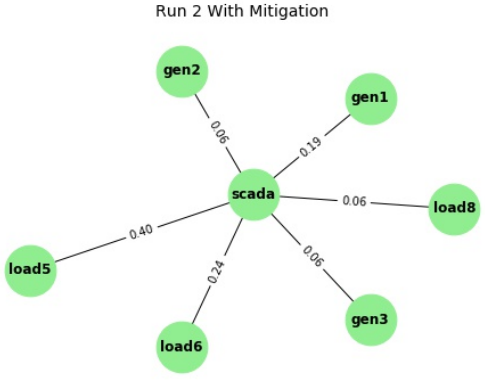
(b) DOS results for Run 3.



(c) No mitigation results for Run 2.

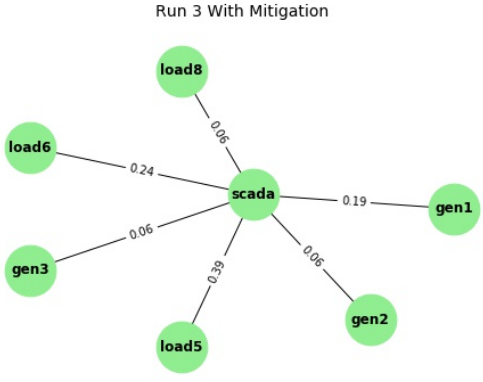


(c) No mitigation results for Run 3.



(d) Mitigation results for Run 2.

Fig. 5: Results for all four scenarios for the second run of the experiments in the dataset. The edges are labeled with the probabilities calculated using the temporal data.



(d) Mitigation results for Run 3.

Fig. 6: Results for all four scenarios for the third run of the experiments in the dataset. The edges are labeled with the probabilities calculated using the temporal data.

observation can be seen where the edges with the highest probabilities are the ones connecting generator 1 (the slack bus) and load 5 to the SCADA node. These results are also justified as this scenario consists of both a cyber-attack and a physical disturbance to the system affecting load shedding. What can be understood from this graph is that the DOS cyber-attack is implemented on load 5 in the network, and the physical disruption affected the load shedding setup in the power system that the SCADA node needed to send more commands to change the generation values at the slack bus (generator 1) to make up for the loss or increase in power generation. These are all valid points to discuss as, once again, the cyber-physical interdependencies can be understood from the dependency graphs (DGs).

Last but not least, observing the With Mitigation results for all runs shows us that the edges with the highest probabilities are the ones connecting loads 5 and 6 to the SCADA node with generator 1 having the second highest probability on the edge connecting it to the SCADA node. These results also make sense as there are firewall rules set up now that prevent the cyber-attack from occurring, hence the increased communication between the SCADA node and the rest of the objects to prevent the attack from occurring.

IV. Conclusions and Future Work

In conclusion, *CyberDep* was developed to generate dependency graphs using the temporal data of the WSCC 9-bus system and perform conditional probability calculations of cyber traffic flows between system nodes. Additionally, we can observe from the results above that the work on *CyberDep* aided in providing insight into cyber-physical interdependencies through quantifying and visualizing the probabilistic relationships between the different system nodes.

Future work includes the consideration of bi-directional traffic flows, integration of more datasets to include additional cyber and physical devices, such as routers and switches, and expansion to larger power systems. *CyberDep* can be utilized to infer and build access paths for cyber-physical threat models and generate cyber-physical kill chains.

V. Acknowledgments

The authors would like to thank Christopher Goes at Sandia National Laboratories for his efforts in generating the datasets used in this work and the members of Sandia Laboratory Directed Research and Development Project #229324 for their collaborative discussions. This work was supported by the Sandia Laboratory Directed Research and Development Project #229324 and the US Department of Energy under award DE-CR0000018.

References

[1] "DNP3 Protocol Primer," DNP Users Group, 2005. [Online]. Available: https://www.dnp.org/Portals/0/AboutUs/DNP3%20Primer%20Rev%20A.pdf

[2] "Cyber attacks on the power grid," IronNet Threat Research, 2022. [Online]. Available: https://www.ironnet.com/blog/cyber-attacks-on-the-power-grid

[3] L. Chen, D. Yue, C. Dou, Z. Cheng, and J. Chen, "Robustness of cyber-physical power systems in cascading failure: Survival of interdependent clusters," *International journal of electrical power & energy systems*, vol. 114, p. 105374, 2020.

[4] K. Marashi, S. S. Sarvestani, and A. R. Hurson, "Identification of interdependencies and prediction of fault propagation for cyber–physical systems," *Reliability Engineering & System Safety*, vol. 215, p. 107787, 2021.

[5] Z. Huang, C. Wang, A. Nayak, and I. Stojmenovic, "Small cluster in cyber physical systems: Network topology, interdependence and cascading failures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2340–2351, 2014.

[6] K. Marashi, S. S. Sarvestani, and A. R. Hurson, "Consideration of cyber-physical interdependencies in reliability modeling of smart grids," *IEEE Transactions on Sustainable Computing*, vol. 3, no. 2, pp. 73–83, 2017.

[7] V. Venkataramanan, A. Srivastava, A. Hahn *et al.*, "Cp-tram: Cyber-physical transmission resiliency assessment metric," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5114–5123, 2020.

[8] A. Clark and S. Zonouz, "Cyber-physical resilience: Definition and assessment metric," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1671–1684, 2017.

[9] A. Sahu and K. Davis, "Structural learning techniques for bayesian attack graphs in cyber physical power systems," in *2021 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2021, pp. 1–6.

[10] P. Żebrowski, A. Couce-Vieira, and A. Mancuso, "A bayesian framework for the analysis and optimal mitigation of cyber threats to cyber-physical systems," *Risk Analysis*, vol. 42, no. 10, pp. 2275–2290, 2022.

[11] S. Krishnamurthy, S. Sarkar, and A. Tewari, "Scalable anomaly detection and isolation in cyber-physical systems using bayesian networks," in *Dynamic Systems and Control Conference*, vol. 46193. American Society of Mechanical Engineers, 2014, p. V002T26A006.

[12] A. AlMajali, Y. Wadhawan, M. S. Saadeh, L. Shalalfeh, and C. Neuman, "Risk assessment of smart grids under cyber-physical attacks using bayesian networks," *International Journal of Electronic Security and Digital Forensics*, vol. 12, no. 4, pp. 357–385, 2020.

[13] X. Lyu, Y. Ding, and S.-H. Yang, "Bayesian network based c2p risk assessment for cyber-physical systems," *IEEE Access*, vol. 8, pp. 88 506–88 517, 2020.

[14] S. Hossain-McKenzie, N. Jacobs, A. Summers, A. Ryan, C. Goes, A. Chatterjee, A. Layton, K. Davis, and H. Huang, "Towards the characterization of cyber-physical system interdependencies in the electric grid," in *2023 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2023, pp. 1–8.

[15] N. Jacobs, S. Hossain-McKenzie, S. Sun, E. Payne, A. Summers, L. Al-Homoud, A. Layton, K. Davis, and C. Goes, "Leveraging graph clustering techniques for cyber-physical system analysis to enhance disturbance characterisation," *IET Cyber-Physical Systems: Theory & Applications*, 2024.

[16] S. Sun, E. Payne, A. Layton, K. Davis, S. Hossain-McKenzie, and N. Jacobs, "Bio-inspired and ai deepwalk based approach to understand cyber-physical interdependencies of power grid infrastructure," in *2023 North American Power Symposium (NAPS)*. IEEE, 2023, pp. 1–6.

[17] S. Hossain-McKenzie, N. Jacobs, A. Summers, B. Kolaczkowski, C. Goes, R. Fasano, Z. Mao, L. Al Homoud, K. Davis, and T. Overbye, "Harmonized automatic relay mitigation of nefarious intentional events (harmonie)-special protection scheme (sps)." Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2022.

[18] "WSCC 9-Bus System," Illinois Center for a Smarter Electric Grid (ICSEG). [Online]. Available: https://icseg.iti.illinois.edu/wscc-9-bus-system/

[19] "SCEPTRE: An Emulation Capability for Industrial Control Systems (ICS)," Sandia National Laboratories. [Online]. Available: https://sandialabs.github.io/sceptre-docs/

[20] H. Li, K. S. Shetye, S. Hossain-McKenzie, K. Davis, and T. J. Overbye, "Investigation of automated corrective actions for special protection schemes," Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Texas A & M ..., Tech. Rep., 2020.

[21] H. Li, K. Shetye, T. Overbye, K. Davis, and S. Hossain-Mckenzie, "Towards the automation of remedial action schemes design," 2021.

[22] C. Vellaithurai, A. Srivastava, S. Zonouz, and R. Berthier, "Cpindex: Cyber-physical vulnerability assessment for power-grid infrastructures," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 566–575, 2014.