REACT to Cyber-Physical Attacks on Power grids (Extended Abstract)*

Saleh Soltan
Electrical Engineering
Princeton University
Princeton, NJ
ssoltan@princeton.edu

Mihalis Yannakakis
Computer Science
Columbia University
New York, NY
mihalis@cs.columbia.edu

Gil Zussman
Electrical Engineering
Columbia University
New York, NY
gil@ee.columbia.edu

ABSTRACT

We study cyber attacks on power grids that affect both the physical infrastructure and the data at the control center– which therefore are cyber-physical in nature. In particular, we assume that an adversary attacks an area by: (i) remotely disconnecting some lines within the attacked area, and (ii) modifying the information received from the attacked area to mask the line failures and hide the attacked area from the control center. For the latter, we consider two types of attacks: (i) *data distortion:* which distorts the data by adding powerful noise to the actual data, and (ii) *data replay:* which replays a locally consistent old data instead of the actual data. We use the DC power flow model and prove that the problem of finding the set of line failures given the phase angles of the nodes outside of the attacked area is strongly NP-hard, even when the attacked area is known. However, we introduce the polynomial time REcurrent Attack Containment and deTection (REACT) Algorithm to approximately detect the attacked area and line failures after a cyber-physical attack.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Reliability, availability, and serviceability; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms, Network problems*

Keywords

Power Grids; Cyber Attacks; Physical Attacks; Information Recovery; Graph Theory; Algorithms

1. SUMMARY

Motivated by the recent cyber attack on the Ukrainian grid [1], in this work, we deploy the DC power flow model and study a model of a cyber-physical attack on the power grid that affects both the physical infrastructure and the data at the control center. We assume that an adversary attacks an area by: (i) disconnecting some lines within the attacked area (by remotely activating the circuit breakers), and (ii) modifying the information (phase angles of the nodes and status of the lines) received from the attacked area to mask the line failures and hide the attacked area from the control center. For the latter, we consider two types of attacks: (i) *data distortion:* which distorts the data by adding powerful noise to the data received from the attacked area, and (ii) *data replay:* which replays a locally consistent old data instead of the actual data. We assume that the system reaches a steady-state after the attack. Fig. 1 shows an example of such an attack.

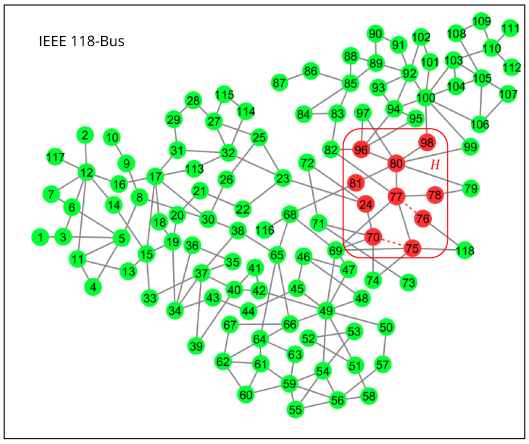


Figure 1: The attack model. An adversary attacks an area H which is unknown to the control center (represented by red nodes) by disconnecting some lines within the attacked area (shown by red dashed lines) and modifying the information received from the attacked area to mask the line failures and hide the attacked area from the control center.

We prove that the problem of finding the set of line failures given the phase angles of the nodes outside of the attacked area is strongly NP-hard, *even when the attacked area is known*. Hence, one cannot expect to develop a polynomial time algorithm that can exactly detect the attacked area and recover the information for all possible attack scenarios. However, we introduce the polynomial time REcurrent Attack Containment and deTection (REACT) Algorithm and numerically show that it performs very well in reasonable scenarios.

The REACT Algorithm combines two modules (named ATAC and LIFD) to provide a comprehensive algorithm for attacked area detection and information recovery following a cyber-physical attack. In particular, the ATtacked Area

*This extended abstract provides a short summary of the paper that appeared in [3] and presented at the Second ACM SIGMETRICS International Workshop on Critical Infrastructure Network Security, University of California, Irvine, June 18, 2018.

Containment (ATAC) Module approximately detects the attacked area using graph theory and the algebraic properties of the DC power flow equations. On the other hand, the randomized LIne Failures Detection (LIFD) Module detects the line failures and recover the phase angles inside the detected attacked area. The LIFD Module builds upon the methods first introduced in [2], to detect line failures using Linear Programming (LP) in more general cases. We prove that in some cases that the methods in [2] fail to detect line failures, the LIFD Module can successfully detect line failures in expected polynomial running time.

We evaluate the performance of the REACT Algorithm by considering two attacked areas, one with 15 nodes and 16 edges (H_1), and the other one with 31 nodes and 41 edges (H_2) within the IEEE 300-bus system. We show that when the attacked area is small, the REACT Algorithm performs equally well after the data distortion and the data replay attacks. In particular, it can exactly detect the attacked area in all the cases, and accurately detect single, double, and triple line failures within the attacked area in more than 80% of the cases (see Fig. 2).

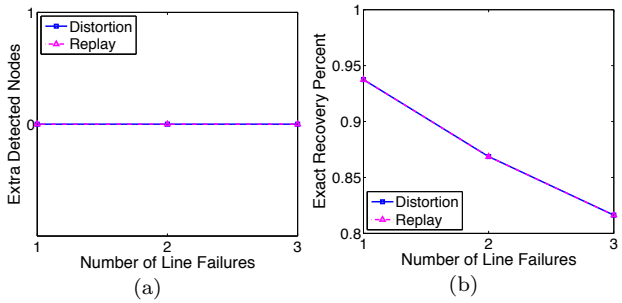
When the attacked area is large, however, the REACT Algorithm's performance is different after the data distortion and the data replay attacks (see Fig. 3). It still performs very well in detecting the attacked area after a data distortion attack and accurately detects line failures after single, double, and triple line failures in more than 60% of the cases. However, it may face difficulties providing an accurate approximation of the attacked area after a replay attack. Despite these difficulties in approximating the attacked area, it accurately detects single and double line failures in around 98% and 60% of the cases, respectively.

The goal of this work is to provide a theoretical foundation for the problem of attacked area and line failures detection after a cyber-physical attack on the power grid. Hence, in this work, we neglect the measurement noise in our analysis and also considered the availability of PMUs at all the nodes. Nevertheless, we demonstrate that this problem is already very challenging without considering these constraints. Extending the results and methods of this paper to the cases where the measurements are noisy and there are limited number of PMUs in the system is part of our future work.

Finally, although the DC power flows only provide an approximation for the more accurate AC power flows, since the ATAC Module for detecting the attacked area mostly depends on the flow conservation checks at each node, the ATAC Module can be easily applied under the AC power flows as well. Moreover, the weight randomization technique and the confidence metric used in the LIFD Module can also be extended to the AC power flows using the methods provided in a recent paper [4]. Extending the results provided in this paper to the transient state of power grids, however, is of particular interest to the power systems community and is part of our future work.

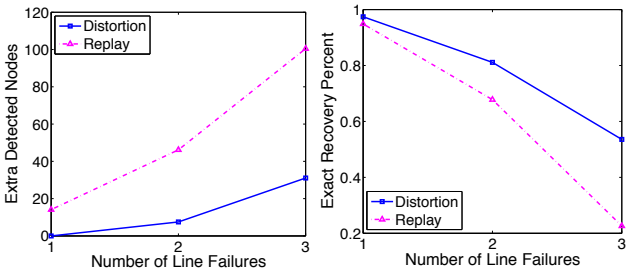
2. ACKNOWLEDGEMENTS

This work was supported in part by DTRA grant HDTRA1-13-1-0021, DARPA RADICS under contract #FA-8750-16-C-0054, funding from the U.S. DOE OE as part of the DOE Grid Modernization Initiative, U.S. DOE under Contract No. DE-AC36-08GO28308 with NREL, and NSF under grant CCF-1703925 and CCF-1423100.



(a) (b)

Figure 2: The REACT Algorithm's performance in detecting the attacked area and recovering the information after data distortion and replay attacks on the attacked area H_1 accompanied by single, double, and triple line failures. (a) Average number of extra nodes detected as attacked in detecting the attacked area, and (b) percentage of the cases with exact line failures detection.



(a) (b)

Figure 3: The REACT Algorithm's performance in detecting the attacked area and recovering the information after data distortion and replay attacks on the attacked area H_2 accompanied by single, double, and triple line failures. (a) Average number of extra nodes detected as attacked in detecting the attacked area, and (b) percentage of the cases with exact line failures detection.

3. REFERENCES

[1] NERC. Analysis of the cyber attack on the Ukrainian power grid, 2016. `http://www.nerc.com/pa/CI/ESISAC/Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf`. Accessed: Jan. 2018.

[2] S. Soltan, M. Yannakakis, and G. Zussman. Joint cyber and physical attacks on power grids: Graph theoretical approaches for information recovery. In *Proc. ACM SIGMETRICS'15*, June 2015.

[3] S. Soltan, M. Yannakakis, and G. Zussman. REACT to cyber attacks on power grids. *to appear in IEEE Trans. Netw. Sci. Eng.*, 2018.

[4] S. Soltan and G. Zussman. EXPOSE the line failures following a cyber-physical attack on the power grid. *to appear in IEEE Trans. Control Netw. Syst.*, 2018.