http://dx.doi.org/10.14236/ewic/ICS2018.7

PSP: A Framework to Allocate Resources to Power Storage Systems under Cyber-Physical Attacks

Yatin Wadhawan
University of Southern California
Los Angeles, CA
ywadhawa@usc.edu

Dr. Clifford Neuman
University of Southern California
Los Angeles, CA
bcn@isi.edu

Dr. Anas AlMajali
The Hashemite University
Zarqa, Jordan
almajali@hu.edu.jo

This risk assessment of the smart grid focuses on energy storage, which is essential but largely unaddressed by the current literature. This work concentrates on actions (such as decreasing or increasing power reserve and power dispatch, performing load curtailment or load shedding, or repair of nodes) the defender should take to meet power demand at minimum operating cost in the presence of cyber-physical attacks on the power and information infrastructure of the smart grid. In this paper, we formulate a Power Storage Protection (PSP) framework against a fixed opponent (adversary). We fix the strategy for the adversary and model the problem as a Partially Observable Markov Decision Process (POMDP) from the perspective of the defender (power utility) and solve it using POMDP solver. We provide a theoretical framework for formulating the above problem and provide experimental results to support our claim using a simplified PSP scenario in which optimal POMDP policy is computed efficiently.

Smart Grid, Distributed Generation, Security Planning, Cyber-Physical Security, Energy Storage Systems

1. INTRODUCTION

Power plants using coal or nuclear fuel take considerable time to start generating power. These conventional centralized plants provide baseload capacity to the electricity grid. To satisfy changes in power demand during peak hours, energy is stored at multiple locations in the smart grid infrastructure. Power utilities incur cost to store energy that can be used during a contingency. When we say energy, we mean fuel that is used to generate power (such as natural gas) or power itself. The energy is stored in different ways: 1) Distributed Energy Resource (DER) is a small-scale power generation and storage unit that use renewable energy such as solar, wind, tidal, etc. DER has the potential to provide efficient power generation and distribution by reducing transmission and distribution losses because they are present near to consumers (Zhu, T., Mishra et al. 2011); 2) natural gas to generate power through Gas Fired Power Plants (GFPP); and 3) power stored in batteries; to meet peak hour demand.

Information and communication technology plays an essential role in supporting distributed power generation, transmission, distribution and load balancing in the smart grid. Advance Metering Infrastructure (AMI) is one example of the use of communication technology that makes the power grid smart. It facilitates communication between utility and customer-end smart meters, providing the knowledge of power rates (Time of use and Real-time pricing) and controlling meters remotely for outage management and demand response. DER units are highly dependent on information technology, which has opened doors for new Cyber-Physical Threats (CPT). The failure of a function in the cyber domain enables the spread of blackouts in North America and Europe (Andersson, G et al. 2003). A cyber attack on the Ukraine power grid (SANS E-ISAC 2016) demonstrated the high likelihood of cyber-physical attacks (CPA) (Neuman, C and Tan 2011). A cyber attack on gas pipeline infrastructure providing gas to GFPP (Wadhawan, Y and Neuman, C 2016); attack on DER units which are responsible for producing, storing and dispatching power, etc. show that cyber attackers have the ability to attack energy storage systems and affect Smart Grid Resilience (SGR). Hence, it is necessary to identify and evaluate the risks associated with smart grid energy storage systems and develop models to perform risk assessment in the presence of various CPAs.

The main motive for utilities is to meet power demand all the times and maintain power grid

stability at minimum cost. To do this they perform the following actions: 1) use power reserves to meet demand, 2) apply load curtailment (Demand Response), and 3) apply load shedding. Load curtailment is an approach used by power utilities to send a request to customers to potentially reduce their electricity usage for a brief period, on demand, in case of power shortage. Load shedding, sometimes called rolling blackouts, is a way to deliberately shut down the power supply in a particular region to prevent failure of the entire power grid. Such actions are taken in response to power shortages due to malicious or non-malicious factors. The main motive of an adversary is to prevent the power utility from meeting power demand, causing cascading blackouts across the region.

An adversary performs a variety of attacks such as topology attacks, integrity attacks and hijacking attacks on the cyber and physical infrastructure of the power grid. The system admin does not know where and what kind of attacks are performed by an adversary. The admin does not know the true state of the system. He receives observations about the change in the system infrastructure such as whether a node or link is Active or Inactive, whether some nodes are malicious or not, etc. The observations are received with some uncertainty from an intrusion detection system installed in the environment.

Most current research efforts are dedicated to different components of the smart grid in the presence of CPAs such as on AMI, DR, AGC, etc. The risk assessment of the power grid focusing on energy storage, which is essential but largely unaddressed by the current literature. Our work concentrates on what actions the defender should take to meet power demand at minimum operating cost in the presence of CPAs by an adversary on the grid's power and information infrastructure.

Our contribution is twofold. First, we describe the different types of attacks on energy storage units and their impact on SGR. Second, we discuss formulation of the Power Storage Protection (PSP) framework against a fixed opponent (adversary). We fix the strategy for the attacker and model the problem as a Partially Observable Markov Decision Process (POMDP) from the perspective of the defender (power utility) (Oliehoek, Frans et al. 2005) and solve it using Increment pruning method (Cassandra, Anthony et al. 1997) using POMDP solver (Anthony R. Casandra 2003-2018). The defender decides what actions to perform in order to minimize the cost of operation and maintain power system stability. The defender decides whether to decrease or increase power reserve, perform load curtailment or load shedding, repair a node or not. The attacker chooses the types of attacks to perform on the infrastructure. In this paper, we provide a theoretical framework for formulating the above problem and provide experimental results to support our claim using a simplified PSP scenario in which optimal POMDP policy is computed efficiently using POMDP solver (Anthony R. Casandra 2003-2018).

The structure of the paper is as follows: Section 2 describes the related work. Section 3 discusses power storage units and the impact of different attacks on them. Section 4 provides a System Description for the PSP. Section 5 describes Experimental and Simulation Analysis. Section 6 and 7 discuss the Conclusion and Future Work respectively.

2. RELATED WORK

The smart grid is dependent on natural gas pipeline distribution to provide fuel to GFPP to generate power during peak hours. Tao et al. (2015) proposed an integrated model for analyzing the impact of the interdependency of natural gas network on power system security. The model incorporates the constraints of the gas network such as daily limits on the pipeline, generating units etc. Using the mathematical model, the paper explains the impact of natural gas supply infrastructure on the economic operation of a vertically integrated utility and discusses the impact of generating unit with fuel switching capability on the power system security when the supply of gas is limited.

Wadhawan, Y et al. (2016) evaluated the resilience of the oil and gas systems in the presence of CPA. The authors demonstrate a function-based methodology and graph-theoretic approach to perform resilience evaluation. Neuman and Tan (2011) described ways that CPAs propagate between the cyber and physical domains. Pan et al. (2016) introduced combined data integrity and availability attacks to expand the attack scenarios against power system state estimation.

Tan et al. (2016) analyzed false data injection attacks on the AGC and provided the understanding of the limits of the physical impact of attacks. Srikantha et al. (2016) proposed a framework that demonstrates the stealthy worst-case strategies for attackers to disrupt the transient stability by controlling DER. Ping et al. (2016) demonstrated the Denial of Service (DoS) attack on the AMI network that reduces the packet delivery rate by 20%.

AlMajali et al. (2013) presented a systems approach to analyze the resilience of the smart grid system in the presence of CPAs. They demonstrate via simulations in Power World how a load altering attack over a period of time destabilizes the power grid. Santos, S. et al. (2017) presented a

mechanism to quantify the impact of energy storage deployment on the level of renewable power integrated into the system. Yan et al. (2016) analyzed the vulnerability of the transmission grid using the Q-learning framework. The authors identify the critical attack sequences by considering the behavior of the physical power system in the presence of sequential topology attacks. They modeled the behavior of the system against different types of attackers.

The concept of game theory has frequently been used in securing cyber-physical systems. Husheng et al. (2012) reduce situational awareness preventing the system admin from responding to attacks immediately. The authors formulated a jamming and anti-jamming game as a zero-sum stochastic game. When remote sensors are jammed, the state information is not delivered to the control system. The actions of the attacker are to decide which sensors to choose to jam. The Nash Equilibrium is computed to demonstrate the increased rewards when effective anti-jamming signals are sent.

Chen, J., & Zhu, Q. (2017) described the significance of microgrid integration in a conventional power system and developed a non-cooperative game theoretic framework to capture the competition and decentralized decisions of the microgrid (act as players). The primary motive of the paper is to develop algorithms that can guide microgrid to take appropriate actions during various events so that overall resilience of the power system is maintained.

Frans et al. (2005) addressed the problem of how to play optimally against the fixed opponent in a two-player card game with partial information. The authors show that if the policy of one player is fixed, we can model the problem as POMDP from the perspective of one agent, and solve using dynamic programming. We are motivated by this approach and model the PSP problem as POMDP where we fix the policy of the attacker. In this way, we compute defender's response against attackers of different categories (naive, nation-state, experienced). In the next section, we discuss different types of attacks on the energy storage and their impact on SGR.

3. ATTACKS ON POWER STORAGE

In this section, we describe possible cyber attacks on energy storage systems and their impact on SGR.

Batteries to store power: The power utilities pay power vendors to reserve power in the batteries. These batteries are connected to the transmission and distribution infrastructure and DER. An adversary performs topology and DoS attacks on the communication infrastructure that is responsible for sending messages for power dispatch. Furthermore, they compromise the communication protocols such as SNMP or Modbus (Alpha Gurdian 2017) to modify the data flowing through the network between power utility and batteries. An attack on batteries will prevent power dispatch during contingency (power shortage to meet demand) and destabilize the grid.

Distributed Energy Resources (DER): DER is smaller power sources that generate, store and dispatch power once signaled from the power utility. They are in a smart building, smart homes, and elsewhere. The need for power during contingency is fulfilled by DERs. Since it is present in consumers facility, consumers can produce power and sell it to the grid. This requires frequent communication between the DER at customer's facility and grid operators. Since they are accessible on the network, they are vulnerable to CPAs. If an adversary is able to control DER, he can disable the system that dispatches the power so that there is a demand-supply mismatch and thus affecting its operation. He performs malware attack remotely on DER, Denial of Service (DoS) attack on the communication infrastructure to prevent power dispatch signals to reach DER, etc.

Natural Gas Storage for Peaker Plants: GFPP generates power during peak hours to meet peak demand. Natural gas is delivered to GFPP via low-pressure distribution pipelines. If an adversary performs an attack on gas pipeline (as described in Wadhawan, Y et al. 2016), this affects the gas delivery to GFPP resulting in loss of power generation during peak hours. The adversary can perform a DoS attack on the communication infrastructure of the gas pipeline system, or a malware attack on wireless nodes. An attack on the natural storage infrastructure will prevent the generation of power during peak hours and destabilize the power grid.

4. POWER STORAGE PROTECTION FRAMEWORK

PSP is an infinite horizon two-player zero-sum Partially Observable Stochastic Game (zs-POSG) with one-sided partial observability. We fix the attack strategy of the attacker and model the problem as a POMDP (Chen, J., & Zhu, Q. 2017) from the perspective of the defender. The main idea behind this approach is that defender does not perform any action if there is no contingency. If the system is working fine, the defender performs NO ACTION.

The model is partially observable because the defender does not know about the true state of the system. He has no idea what attacks happen. The game is zero-sum because attacker wants to

perform attacks that destabilize the power grid and maximize defender's cost. For each attacker action, defender performs an action which incurs some cost. And defender wants to decide which action to take at every time step that stabilizes the grid and minimizes the cost.

PSP POMDP model is defined by a tuple Γ = *(I, S, B, A, T, R, b^0, Ω, O, Ψ,* γ*)*. Nomenclature:

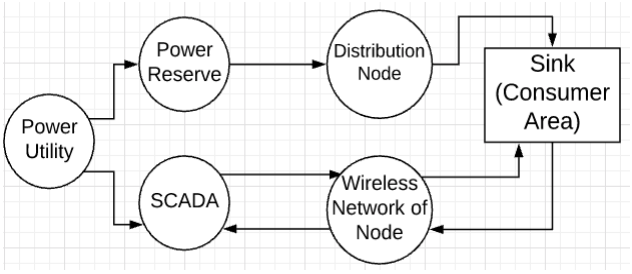

Figure 1: *System State*

- $I = \{I^d, I^a\}$ is the set of agents.
- S is the finite set of world states.
- $B: \Delta(S)$ is the probability distribution over S.
- A is the finite set of actions (A^d, A^a) of the agents.
- $b^0 \in \Delta(S)$: An initial belief of the game.
- Ω is the finite set of observations.
- O is the conditional observation probabilities.
- Ψ is the fixed stochastic strategy of the attacker.
- T: $S \times A \rightarrow S$ is the state transition function.
- R is reward function R: $S \times A \rightarrow \Re$ for agent I^d.
- $\gamma \in [0,1]$ is a discount factor.

Suppose $s \in S$ is the current state of the system at time step t. In s, the defender takes action $a_i \in A^d$, and the attacker takes action $a_j \in A^a$ according to the fixed stochastic policy $\Psi_j = p(a_j \mid s, a_i)$. The game moves to new state $s' \in S$ according to a stochastic joint transition model $p(s' \mid s, a_i, a_j)$ in time step t+1. Since we know the policy of the attacker (fixed), we compute a single transition model $T(s' \mid s, a_i)$ for the defender:

$$p(s' \mid s, a_i) = \Sigma_{a_j} p(s' \mid s, a_i, a_j)\, p(a_j \mid s, a_i) \quad (1)$$

Using this equation, we model the game as POMDP from the perspective of the defender (Frans et al. 2005). The defender receives an observation $o \in \Omega$ with probability $O(o \mid s', a_i) = p(o \mid s', a_i)$ as the game moves to a new state s'. Since the strategy of the attacker is fixed, we do not care about the attacker's observation. The defender receives a reward (cost) of $-R_t(s, a_i)$ for this transition and the attacker receives $R_t(s, a_i)$ because this is a zero-sum game. The reward may depend on the previous state of the game, and the joint action (eq. 2). It is also possible that rewards just depend on the state of the system (eq. 3).

$$R_t(s, a_i) = \Sigma_{a_j} r(s, a_i, a_j)\, p(a_j \mid s, a_i) \quad (2)$$

$$R_t(s, a_i) = r(s) \quad (3)$$

The key assumption of POMDP is that the world states (S) are not fully observable, and therefore the concept of belief state is introduced (which is the probability distribution over world states S). That is how we transform POMDP into Belief-MDP where transition, observation and reward functions are over belief space. Initially, the defender has an initial belief of b^0. The belief gets updated at every time step based on the action and observation pair:

$$b(s') = \eta\, p(o \mid s', a_i) \Sigma_s p(s' \mid s, a_i)\, b(s) \quad (4)$$

$$\eta = 1 / p(o \mid b, a_i) \quad (5)$$

$$p(o \mid b, a_i) = \Sigma_{s'} p(o \mid s', a_i) \Sigma_s p(s' \mid s, a_i)\, b(s) \quad (6)$$

where η is the normalizing constant. The transition function from belief state b to b' when defender takes action $a_i \in A^d$:

$$T(b, a_i, b') = \Sigma_o p(b' \mid b, a_i, o)\, p(o \mid b, a_i) \quad (7)$$

where $p(b' \mid b, a_i, o) = 1$ if belief update with arguments b, a_i, o returns b', otherwise 0. And the reward function of taking action $a_i \in A^d$ in belief state b:

$$R(b, a_i) = \Sigma_s R_t(s, a_i)\, b(s) \quad (8)$$

The main goal of PSP is to find the sequence actions (for corresponding attacker's actions) that maximize the expected rewards for the defender for each belief. The value function for each belief is represented by eq. 9. We will discuss how to solve PSP POMDP in section 5. In the following subsections, we describe the domain and problem statement in detail.

$$V(b) = \max_{a_i \in A^d} [\, R(b, a_i)] + \gamma \Sigma_{b'} T(b, a_i, b')\, V(b') \,] \quad (9)$$

4.1 Agents

There are two agents (I) in this game. One is an adversary I^a (a cyber hacker), and second is defender I^d (power utility). The adversary may be an insider (who wants to take revenge), state-sponsored, or terrorist hackers. The main motive is to reduce the resilience of the power grid by preventing the power utility to meet the power demand. The main motive of the defender is to meet the power demand all the time at minimum cost

4.2 System State

The state *(S)* of the system is represented in the form of two directed graphs. Graph G1: represents the power distribution network (see Fig. 1 upper portion). The nodes are power reserve nodes, power distribution nodes, and power sink nodes.

The power reserve (PR) nodes store a certain amount of power to meet unexpected demand. The power utilities (PU) pay some cost to maintain PR. PR nodes are connected to power distribution (PD) nodes, and finally, PD nodes are connected to client (C) nodes, which consume power. Here, we abstract a particular zip code power demand in one client node. The edges between these nodes represent the power flow from PR to PD to C. Each edge (E^P) has power capacity (C_e) and amount of power actually flowing through it (U_e). The conservation of power flow is followed by each node and edge in the graph. The amount of power going inside the PD node is equal to the amount of power going out of it. The power flow through all edges follows: $U_e \leq C_e$. The client nodes have certain power demand (always consume power, edges go into them) that may change over time. The power reserve nodes are power sources that always produce (edges go out of them).

Graph G2: represents the information flow between the client nodes and PU (see Fig. 1 lower portion). How a PU figures out demand in a particular region? Based on the AMI, the information about client power consumption and demand is sent to the PU. On the basis of readings, the PU decides what action to perform to meet the demand and maintain system resilience. We have three types of nodes in the graph. One is PU to gather information, second is routers, intermediate nodes to transmit information, and finally client nodes to send power reading. Over this network, PU sends commands to clients to perform load shedding, curtailment, etc. The edges (E^I) represent the information flow from PU to C and vice versa. The edge in the information network can either be *ACTIVE* or *INACTIVE* and nodes in the information network can either be *MALICIOUS* or *NON-MALICIOUS* based on the type of attack performed by an adversary.

4.3 Actions

From a defender's point of view, there are two categories of actions: Cyber Actions and Physical Actions. The cyber actions are performed from the cyber domain. It is further divided into two categories: Cyber-Cyber and Cyber-Physical Protection. The cyber-cyber actions tend to protect the cyber components from cyber-attacks. For instance, vulnerability assessment and patching vulnerabilities prevent a system from getting compromised by an adversary. The cyber-physical protection actions are taken from the cyber domain on the physical infrastructure so that the power grid continues to meet the power demand. For instance, load curtailment or load shedding commands are sent from the cyber domain via demand response mechanism to reduce the amount of power consumption at the customer end.

The physical actions consist of actions that are performed in the physical domain. For instance, repairing a compromised node, link in the infrastructure, manually updating the software of a Programmable Logic Controller (PLC), etc. Physical actions are costlier than cyber actions because the defender must send a technician to repair parts of the network and it also takes time. In this paper, we are concerned about the cyber-physical and physical actions of the defender. We do not consider the cyber-cyber actions (such as Patching or scanning). The action set for the defender is:

A^D *: {Cyber-Physical Actions: No-Action, Increase Power Reserve, Reduce Power Reserve, Load Curtailment, Load Shedding; Physical Actions: Repair Node, Repair Link}.*

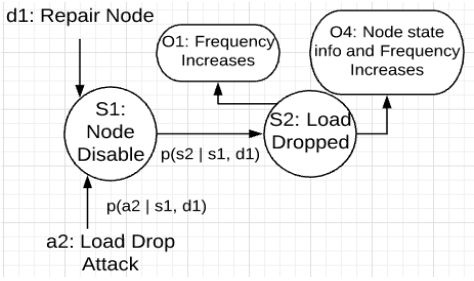

Figure 2: *State Transition*

We have also included No-Action for the defender in case the system is in a consistent state with no action performed. The adversary acts by performing attacks on the information network. The adversary can perform a variety of attacks such as malware injection in PLCs or AMI meters, data integrity, etc. with a motive to prevent correct power demand information reaching to the defender. There are three categories of attacks that can be performed:

A^A *: {Topology attacks, Integrity attacks and Hijacking}.*

The topology attacks include removal of a node or link in the information network. It prevents power consumption information from reaching the utility. Integrity attacks alter the readings from the node (meters). This is much harder to detect and hence a defender may be deceived. The integrity attack is performed by hijacking a node or link where the attacker issues fake commands of actions available to the defender such as load curtailment. Each action performed by the agents has some cost involved and provides a reward. We will discuss the reward and cost structure in the following sections.

4.4 State Transition and Observations

The state transition is determined by the choice of the defender actions (which exhibit a deterministic response) and attacker actions (which also exhibit

a deterministic response but occur stochastically). The state of the system is represented in terms of two graphs G1 (power graph) and G2 (information graph). The agents perform actions (described in the previous subsection). The defender performs action first in a particular belief followed by the attacker's action. The system moves to a new belief state and defender receives some observation. The following observations could be received by the defender:

O^D: {Node failure, Link failure, Node and Link failure, Area Frequency goes up, Area Frequency goes down, Attack on AGC, Attack on DER, Attack on AMI}.

Consider Fig. 2 for state transition example. Suppose system is in state s1: Node Disable, where attacker has already disabled the wireless node in the communication network. The main motive of the defender is to repair this node to maintain state awareness and send signals to consumers. The defender performs action d1: Repair node. The attacker performs action a2: Load Drop attack with some probability p(a2 | s1, d1). Due to this, the frequency increases. Note: normally, the area frequency limit is 60 Hz. The frequency protection is enabled with a threshold of 62.4 Hz (1.04 pu) and under frequency is 57.60 Hz (0.96 pu ratio) with a pickup time of 2 seconds. If the frequency in the system exceeds 62.4 Hz for more than 2 seconds, generators will trip in response to over frequency protection mechanism and same in under frequency case. The over frequency happens when there is more generation than load and under frequency when there is lower generation than load.

The system moves to a new state s2: Load Dropped. The defender will observe either of two observations: o1: Frequency increase or o2: Node Disable and Frequency increase. And finally, defender incurs the cost because of load drop attack. On the basis of the observation, the defender performs an action and system continues to move in the new state. In section 5, we describe this example in detail for the purpose of experiment.

4.5 Payoffs

We define the payoff of the defender at every time step t. The defender problem is the multi-objective optimization problem. The defender needs to stabilize the grid by taking actions to fix attacks at minimum cost. The defender payoffs depend on the number of the following factors.

Distance from the ground truth Power Demand: The main function of the defender is to meet power demand at each time step with minimum operating cost. Suppose the minimum power storage is D is always maintained by the defender. There always exist a ground truth power demand at each time step t, pd_t, which defines the total amount of power required to the customers c:

$$pd_t = \Sigma_{c,\,t}\ \mathrm{pd}_c \tag{10}$$

where $c \in C$ is the number of customers nodes present in the power network G1. The defender would like to provide the ground truth power demand at the lowest cost possible. The distance from the above ground truth at time t is calculated by taking its difference from the total amount of power reserve available with the defender.

$$pr_t = \Sigma_{r,\,t}\ \mathrm{pr}_c \tag{11}$$

where $r \in R$ is the number of power reserve nodes present in the power network. The difference at time t is pr_t - pd_t. The positive difference depicts excess power storage and negative difference means there is shortage of power shortage. The main purpose of the defender is to minimize the mod value of the above difference, i.e.,

$$| pr_t - pd_t | \leq \varepsilon \tag{12}$$

$\varepsilon \rightarrow \Re$ is threshold value up to which power frequency is maintained in the grid and it is stable. If power shortage crosses this threshold, generators will trip due to under frequency protection mechanism. If it is maintained, the defender will incur a cost of storing more power. We include a term $C_{per\ unit}$, which denotes the cost of reserve per unit. It is positive when there is excess power, otherwise zero. Therefore, the distance from the ground truth is defined by $| pr_t - pd_t | + C_{per\ unit}$.

Cost of Repairing Action (αC_R): If defender performs repairing of an INACTIVE or MALICIOUS node, there is inherent cost involved with it. α is Boolean variable to decide whether repair action has been taken or not. And C_R is the fixed cost for repairing.

Number of Inactive or Malicious links (N_{IM}): The payoff for defender also depends on the percentage of nodes and link that are MALICIOUS and INACTIVE.

Payoff defender is represented as D_t:

$$D_t = \mathrm{int}\ (\ | \mathrm{pr}_t - \mathrm{pd}_t | + C_{\mathrm{per\ unit}} + \alpha C_R + N_{IM}) \tag{13}$$

where int represents the normalization of the value. For instance, distance is divided by the maximum distance. The main motive of the defender is to take actions so to minimize D_t at every time step given the fixed policy of the attacker.

Note, when we model the problem in the form of POMDP we specify the payoffs. The question that arises is how to know the payoffs for different states before it occurs because payoffs depend on the factors that are determined at each time step. We compute payoffs in the following way. First, for

the number of nodes get disabled, the states are different. For example, s is a state when one node is disabled and s' is one where two nodes are disabled. And we know the cost when we know the number of nodes disabled in a state. Second, the cost of repairing a disabled node is standard specified by the defender.

Finally, the cost of storing power reserve depends on the demand and reserve at each time step. The gap cannot be more than ε otherwise, power system will destabilize because power frequency crosses under or over protection threshold. So we do not compute the value of gap more than equal to ε. For values less than ε, the defender maintains D amount of power reserve always. For the difference from the D, we will assign average cost the defender has incurred in the past. This will simplify the model generation.

4.6 Assumptions

We have taken following assumptions while formulating the PSP POMDP model.

- We do not have real world data so we randomly assign probabilities for state transition, observations, attacker's policy and payoff values in our simulation. The probability of transition $p(s' \mid s, a_i)$ is computed based on eq. 1 after assigning the probabilities for $p(s' \mid s, a_i, a_j)$ and $p(a_j \mid s, a_i)$.
- The actions performed by the defender and attacker have deterministic response.
- The amount of time it takes for an action to perform and take effect is not considered.
- It is difficult to scale the problem if you do not have real-world data because we have to assign the probability for each transition of the system, agents actions, and rewards. The future work is to find out ways to scale the problem with and without data and use above payoff method to compute payoffs.
- The power utility may have contracts with many power reserve companies and they charge differently for power at different times of day. We have not considered this scenario in this paper.

Table 1: *List of Defender Actions in the test network.*

```
Defender:
actions: d0 d1 d2
d0: nothing
d1: node-repair
d2: reduce-power-reserve-dispatch
```

Table 2: *List of Attacker Actions in the test network.*

```
Attacker:
actions: a0 a1
a0: node-disable
a1: load drop attack
```

Table 3: *List of states in the test network.*

```
states: s0 s1 s2 st
s0: normal
s1: node-removed
s2: load-drop
st: node-removed-load-drop
```

Table 4: *List of observations in the test network.*

```
o1: frequency-increases
o2: no-state-info
o3: normal-scenario
o4: no-state-info-frequency-increases
```

Table 5: *Observation probabilities in test network.*

```
O: * : s0 : o3      O: * : s2 : o4
1.000000            0.100000
O: * : s1 : o2      O: * : st : o1
0.900000            0.100000
O: * : s1 : o4      O: * : st : o2
0.100000            0.100000
O: * : s2 : o1      O: * : st : o4
0.900000            0.800000
```

Table 6: *Rewards corresponding to states in the test network.*

```
R: * : * : s0 : *
-10
R: * : * : s1 : *
20
R: * : * : s2 : *
10
R: * : * : st : *
30
```

Table 7: *Transitions for action d0, d1, and d2.*

```
T: d0 : s0 : s0   T: d1 : s0 : s1   T: d2 : st : s2
1.00000           0.400000          0.400000
T: d0 : s1 : s1   T: d1 : s0 : st   T: d2 : s0 : s1
1.00000           0.200000          0.500000
T: d0 : s2 : s2   T: d1 : s1 : s1   T: d2 : s1 : s1
1.00000           0.500000          0.500000
T: d0 : st : st   T: d1 : s2 : st   T: d2 : s2 : s1
1.000000          0.400000          0.400000
                  T: d1 : s2 : s2   T: d2 : s2 : s0
                  0.600000          0.450000
                  T: d1 : st : st   T: d2 : st : s1
                  0.350000          0.400000
                  T: d1 : st : s2   T: d2 : st : st
                  0.650000          0.200000
                  T: d1 : s0 : s2   T: d2 : s0 : s2
                  0.400000          0.500000
                  T: d1 : s1 : s2   T: d2 : s1 : st
                  0.400000          0.500000
                  T: d1 : s1 : s0   T: d2 : s2 : s2
                  0.100000          0.150000
```

Table 8: *POMDP Solver Policy Graph. N stands for Node Id, A for Action, o for observations.*

N	A	o1	o2	o3	o4	N	A	o1	o2	o3	o4
0	1	14	0	16	15	9	1	12	8	16	5
1	1	14	2	16	15	10	2	15	3	16	4
2	1	14	4	16	15	11	2	14	3	16	5
3	1	12	8	16	10	12	2	12	7	16	3
4	1	12	3	16	13	13	2	15	8	16	4
5	1	14	3	16	13	14	2	12	7	16	6
6	1	12	3	16	10	15	2	12	8	16	4
7	1	12	7	16	6	16	0	12	0	16	12
8	1	12	8	16	4						

5. EXPERIMENT

In this section, we provide details of the system we have considered for the simulation and discuss results.

5.1 POMDP Model

We have generated a POMDP model manually to demonstrate the concept in this paper. Generating a POMDP model for power demand satisfying game requires knowledge about possible states, actions, observations and rewards agents receive. In real life, the actions and observations can be derived from the experience of the defender (system admin) and using tools such as IDS. The states of the system can be derived from the history of the system by determining what the different states of the system were in the past. The admin's experience should be used to assign cost incurred to take action to protect the system. The initial belief of the system would start from the normal state because the defender will perform no action if there is no attack.

Consider Fig. 1 as a test network for the simulation. We have a SCADA node to monitor the state of the system and take actions, customer node (abstracts a zip code) to send information about power consumption and receives commands from the power utility, power distribution node (intermediate node) and power reserve node that stores power in the form of batteries.

We define the POMDP model in the form POMDP file as described in (Anthony R. Casandra 2003-2018). The actions of the defender and attacker are defined in Table. 1 and 2 respectively. The probability of an attacker taking a particular action depends on the state and action of the defender. For simulations, we assume that the policy Ψ has probability equal to the fraction of the number of actions available to the attacker. A naive attacker does not know what best action to take in a state. For him all actions are equal. Since there are two actions available to the attacker the probability is 0.5.

Table 3 describes the list of states of the system possible if attacks happen on the system shown in Fig. 1. For instance, s1 is a node removed state when attacker performs physical action: node disable. Table 4 describes the list of observations defender receives when the system moves to a new state. Table 5 list the observations received at a state with probabilities. The observations depend on the state of the system. Table 6 list the cost defender receives in a particular state. The rewards are dependent on the state of the system. Table 7 represents the transition from a state to another state on the basis of the actions taken by the defender (one column for each action). The probability of transition is calculated by considering the possible actions taken by the attacker in a particular state according to eq. 1. The value of $p(a_j \mid s, a_i)$ (attacker's probability to take action a_j) is 0.5 for all defender actions and state.

5.2 Solving POMDP Model

We use POMDP solver (Anthony R. Casandra 2003-2018) (written C) to compute optimal policy for the defender against a fixed attacker. The solver uses basic dynamic programming approach for the algorithms, solving one stage at a time. It will stop solving if the answer is within a tolerable range of the infinite horizon. The POMDP solver takes input a POMDP file of the format shown in Table 1 to 7. It computes the optimal value function vector coefficients and optimal policy graph (in Table 8) on the basis of the observation received by the agent using Incremental pruning algorithm (Cassandra, Anthon et al. 1997). The simulation runs on machine with Intel Core i5 at 2.4 GHz and 8 GB RAM. The solver is run without time horizon limit and with discount factor of 0.95. The total time it takes to solve the POMDP model is 13.31 secs.

5.3 Simulation Analysis

The simulation results is in the form of value function vectors and policy graph. Each line of the policy graph (in Table 8) represents one node with unique node ID (N). It is numbered sequentially and lining up sequentially with the value function vectors file. The node ID is followed by action number (A), which is further followed by a list of node IDs, one for each observation (o). This list specifies the transitions in the policy graph. The o'th number in the list will be the index of the node that follows this one when the observation received is 'o'.

As an illustration of the optimal policies found by the POMDP solver, consider a simple case where the defender is in some belief state b (say node ID 16 in Table 8). In belief state b, he observes o3 where everything is working fine. POMDP solution recommends him to jump to node ID 16 and perform action d0 (No-Action). So he remains in the same belief state. If he observes o1 (frequency increases) in b, solution recommends to jump to node ID 12 and perform action d2 (reduce power reserve dispatch) so that to stabilize the frequency of the system with threshold limits. And if defender observes o4 (no state info and frequency decrease) in b, solver recommends d2. Note, in all node IDs when defender observes o4, half of the time solution recommends d1 and d2 for another half. It is because o4 means node is disabled and there is load drop attack. In all node IDs, when defender observes o3 (normal scenario), the belief state jumps to 16 to perform d0 (No-Action). This shows that POMDP model is able to take an effective

decision that will maintain system resilience and minimize the cost of the defender.

According to assumptions, in this experiment we have not assigned ε threshold value. In reality, if the gap between power demand and power reserve is more than ε, the solver will recommend the same action that is to reduce the power reserve dispatch. Lets take another scenario where power demand increases and frequency goes down, the defender either increase the power reserve dispatch, perform load shedding or load curtailment. If this scenario is given to the POMDP solver, the solver will recommend an action that will stabilize the grid and minimize the cost. We have to specify the actions load shedding and load curtailment and cost of performing these actions in the model.

6. CONCLUSION

In this paper, we formulated the Power Storage Protection game against a fixed naive adversary. We fix the strategy for the attacker and model the problem as a POMDP from the perspective of the defender (power utility) and solve it using POMDP solver. We provide the theoretical framework for formulating the PSP problem and provide experimental results to support our claim using a simplified PSP game in which optimal POMDP policy is computed efficiently. Our experimental results show that defender can compute the optimal policy against a fixed attacker policy. Through this formulation, the defender can learn the optimal policy against different classes of attackers. The challenge is to compute all the probabilities for transition, observation, and rewards for each state to generate the POMDP model. The system admin should compute these values by carefully considering various factors discussed in this paper.

7. FUTURE WORK

We plan to develop techniques to scale the problem both with and without real world data. We plan to use the payoff method (described in Section 4.5) to compute payoffs for the POMDP model. Furthermore, we will consider alternative fixed strategies Ψ for different categories of the attacker and simulate the defender's response against them.

8. REFERENCES

Li, Tao, Mircea Eremia, and Mohammad Shahidehpour. "Interdependency of natural gas network and power system security." *IEEE Transactions on* Power Systems 23, no. 4 (2008): 1817-1824.

Wadhawan, Y., & Neuman, C. (2016, October). Evaluating Resilience of Gas Pipeline Systems Under Cyber-Physical Attacks: A Function-Based Methodology. In *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy* (pp. 71-80).

Neuman, C., and Tan, K.: Mediating cyber and physical threat propagation in secure smart grid architectures. In Smart Grid Communications (SmartGridComm), IEEE International Conference on (pp. 238-243). (2011)

Tan, Rui, Hoang Hai Nguyen, Eddy YS Foo, Xinshu Dong, David KY Yau, Zbigniew Kalbarczyk, Ravishankar K. Iyer, and Hoay Beng Gooi. "Optimal False Data Injection Attack against Automatic Generation Control in Power Grids." In 2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS), pp. 1-10. IEEE, 2016.

Srikantha, Pirathayini, and Deepa Kundur. "A DER Attack-Mitigation Differential Game for Smart Grid Security Analysis." IEEE Transactions on Smart Grid 7, no. 3 (2016): 1476-1485.

SANS E-ISAC. 2016. Analysis of the Cyber Attack on the Ukrainian Power Grid, March 2016. http://www.nerc.com/pa/CI/ESISAC/Documents/ E- ISAC_SANS_Ukraine_DUC_18Mar2016.pdf

Yi, Ping, Ting Zhu, Qingquan Zhang, Yue Wu, and Li Pan. "Puppet attack: A denial of service attack in advanced metering infrastructure network." Journal of Network and Computer Applications 59 (2016): 325-332.

AlMajali, A., Rice, E., Viswanathan, A., Tan, K., & Neuman, C. A systems approach to analysing cyber-physical threats in the Smart Grid. In 2013 IEEE International Conference on Smart Grid Communications.

Zhu, T., Mishra, A., Irwin, D., Sharma, N., Shenoy, P., & Towsley, D. (2011, November). The case for efficient renewable energy management in smart homes. In Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (pp. 67-72). ACM.

Santos, S. F., Fitiwi, D. Z., Cruz, M. R., Cabrita, C. M., & Catalão, J. P. (2017). Impacts of optimal energy storage deployment and network reconfiguration on renewable integration level in distribution systems. Applied Energy, 185, 44-55.

Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., Kundur, P., ... & Schulz, R. (2005). Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. IEEE Transactions on Power Systems, 20(4), 1922-1928.

Pan, K., Teixeira, A. M., Cvetkovic, M., & Palensky, P. (2016, November). Combined data integrity and availability attacks on state estimation in cyber-physical power grids. In Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on (pp. 271-277). IEEE.

Yan, Jun, Haibo He, Xiangnan Zhong, and Yufei Tang. "Q-learningbased vulnerability analysis of smart grid against sequential topology attacks." IEEE Transactions on Information Forensics and Security 12, no. 1 (2017): 200-210.

Li, H., Lai, L., & Qiu, R. C. 2011. A denial-of-service jamming game for remote state monitoring in smart grid. In Information Sciences and Systems (CISS), 2011 45th Annual Conference on (pp. 1-6). IEEE.

Chen, J., & Zhu, Q. (2017). A game-theoretic framework for resilient and distributed generation control of renewable energies in microgrids. IEEE Transactions on Smart Grid, 8(1), 285-295.

Cassandra, Anthony, Michael L. Littman, and Nevin L. Zhang. "Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes." In Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence, pp. 54-61. Morgan Kaufmann Publishers Inc., 1997.

Oliehoek, Frans, Matthijs TJ Spaan, and Nikos Vlassis. "Best-response play in partially observable card games." In Proceedings of the 14th annual machine learning conference of Belgium and the Netherlands, pp. 45-50. 2005.

Alpha Guardian. 2017. Energy Storage System Vulnerabilities.http://www.alphaguardian.net/energy-storage-system-cyber-vulnerabilities/

Anthony R. Casandra 2003-2018. Partially Observable Markov Decision Process. http://www.pomdp.org