SCORE: A Security-Oriented Cyber-Physical Optimal Response Engine

Abhijeet Sahu
Electrical Engineering
TAMU, College Station
abhijeet_ntpc@tamu.edu

Hao Huang
Electrical Engineering
TAMU, College Station
hao_huang@tamu.edu

Katherine Davis
Electrical Engineering
TAMU, College Station
katedavis@tamu.edu

Saman Zonouz
Electrical Engineering
Rutgers University
saman.zonouz@rutgers.edu

***Abstract*—Automatic optimal response systems are essential for preserving power system resilience and ensuring faster recovery from emergency under cyber compromise. Numerous research works have developed such response engine for cyber and physical system recovery separately. In this paper, we propose a novel cyber-physical decision support system, SCORE, that computes optimal actions considering pure and hybrid cyber-physical states, using Markov Decision Process (MDP). Such an automatic decision making engine can assist power system operators and network administrators to make a faster response to prevent cascading failures and attack escalation respectively. The hybrid nature of the engine makes the reward and state transition model of the MDP unique. Value iteration and policy iteration techniques are used to compute the optimal actions. Tests are performed on three and five substation power systems to recover from attacks that compromise relays to cause transmission line overflow. The paper also analyses the impact of reward and state transition model on computation. Corresponding results verify the efficacy of the proposed engine.**

***Index Terms*—MDP, cyber-physical systems, value and policy iteration, optimal response**

I. Introduction

The frequency of cyber attacks directed toward power systems is increasing, and such attacks are becoming increasingly complex. Growing interdependencies between information systems and power networks contribute to the need to represent and defend power grids as vulnerable cyber-physical systems [1]. Intelligently crafted malicious attacks have the potential to cause global system failure. Industrial Control Systems (ICS) include, for example, Supervisory Control and Data Acquisition (SCADA) for power system Energy Management Systems (EMS); these are major components of modern power grids that provide multiple functionality ranging from state estimation, to optimal power flow, to economic dispatch, etc [2, 3, 4]. These ICSs that control critical power infrastructure are vulnerable to diverse cyber attacks such as side-channel attacks, network intrusions, false data injections attacks, and other attacks over the Internet Protocol (IP) network. Discovery of such cyber intrusions can create an immediate panic situation in the control rooms which may complicate the operator's task to determine the immediate recovery action. To ensure grid resilience to such attacks and to avail power supply to loads during such a compromised state, an automatic optimal response engine can assist an operator to take an effective remedial action as well as guide a network administrator to address cyber attack spread. Such a response system needs to consider the state of the system and the available resources to restore the system back to a previous good state or to release the stress. The challenge is that the state representation cannot be confined to pure physical states. For instance, a Man-in-the-Middle attack performing data manipulation of an over-current relay protecting a transformer, may obfuscate an operator to trip a breaker unnecessarily and disrupt the normal operation.

Moreover, the dynamic power network topology and state makes the decision making arduous for the operator when predicting the root cause of the contingency; whether caused by cyber intrusion or system faults. Sometimes the operator may be posed with multiple control actions to decide upon. To address this difficulty, ideas from artificial intelligence for automatic optimal response can be leveraged. Proposals from soft computing fraternity such as fuzzy systems [5], artificial neural networks [6], evolutionary computing such as genetic programming and algorithm [7], ant colony optimization [8], particle swarm optimization [9], simulated annealing [10], probabilistic reasoning [11] etc. can be utilized in solving diverse kinds of optimal response problems, depending on availability of labeled data, computational capacity and response time constraints. Hence, automatic response engines can potentially help power system operators make better decisions, particularly under complex cyber attack and other threat scenarios.

In this paper, we propose an MDP-based cyber-physical optimal response engine with hybrid states and actions from both cyber and physical domains. Our main contributions are in, a) modeling the reward function and state transition model for the MDP; b) analyzing the impacts of the MDP components on the computation and accuracy of the optimal response; and c) testing our engine using value and policy iteration MDP solver for three and five substation power system cases.

The paper proceeds as follows. Section II provides the literature review on models created for optimal response problems. Section III introduces fundamentals of MDP, value and policy iteration techniques. In Section IV, the details of our threat, fault and optimal response model are described. Results and analysis on the proposed model is successfully tested for 3 and 5 substation power system in Section V. Finally, we conclude the paper with the scope of future work and conclusion in Section VI.

II. Literature Review

As a cyber-physical system, it is necessary to consider power systems from both cyber and physical domains for providing an optimal response for overall power system security and resilience. [3, 12, 13] explored different approaches to identify and rank the possible contingencies in power systems by considering cyber vulnerabilities that could be induced by adversaries along with the resulting physical grid attack impact to provide cyber-physical risk analysis and situational awareness for impending contingencies. As per NERC [14],

Accepted for publication in IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, October 2019

a contingency is an event that can occur in the future, such as an outage of a generator or circuit breaker, that need to be dealt with and must be prepared for. Based on that, this paper extends the cyber-physical model with the fusion of cyber and physical vulnerabilities in power systems for optimal contingency response.

Taking the relay as an example, Fig. 1. presents the proposed fused cyber-physical model. To simplify the model, this paper uses cyber and physical attacks to describe a few possible attack scenarios in cyber and physical networks. In [15], Liu et al presented and analyzed the impact of different attacks, including untimely data, communication outage, Denial of Service, etc., on power grids with their cyber-physical testbed. In [16], Hong et al presented different intrusion scenarios considering the substation level topology in power systems. For relays, compromise can occur by either cyber or physical attack and lead to undesired behaviors, including lockout, tripping, and delayed operation. Such adversaries could cause the system to lose the load, lose stability, or have an unexpected overflow [15, 16, 17]. With different adversary scenarios, the relay can be compromised into following states: *Relay Lockout*, *Relay Open*, and *Relay Delay Operation*. Then, the system falls into *Loss of Load*, *Power System Overflow*, or *Power System Transient Instability*. Our current work only considers the scenario of *Relay Open* in the Physical Action layer and *Power System Overflow* in the System Reaction layer as the consequences. In the future, we will expand our engine to incorporate other compromises and action scenarios with a more comprehensive model.

Different Markov Models have been proposed for cyber-physical modeling. An MDP is a type of Markov model that can be applied for optimal decision making. For a pure cyber system, in [18], an MDP based automatic Intrusion Response System (IRS) is proposed that selects optimal long-term responses from atomic response actions to protect the cyber components, reducing threat resolution time. A network-based attack mitigation technique is proposed by building a response selection model evaluating the negative and the positive impacts of the actions [19]. Considering the pure physical system, an automatic generation control (AGC) problem is modeled as a stochastic multistage decision problem [4]. From the attacker's perspective, an Reinforcement Learning (RL) based False Data Injection attack has been proposed to disrupt Automatic Voltage Control [2]. Authors in [20] propose a Q-learning-based vulnerability analysis to prevent Q-learning-based sequential topology attacks to cause cascading failure in the smart grid. One of our previous works developed a response and recovery engine using an Attack Response Tree (ART), where the optimal responses were obtained by solving a Partially Observable MDP derived from the automatically generated ARTs [21]. The state space in this model was confined to the cyber side only. In our current work, we propose a mixed MDP model that constitutes states and actions from both domains to recover the power system from emergency.

III. Background

A. Basics of Markov Decision Process

A Markov Decision Process is a discrete-time stochastic process used to describe the agent and environment interactions. In our problem, the agents are compared to the decision making engines and the environment is the current state of the cyber-physical system. It is defined by the tuple

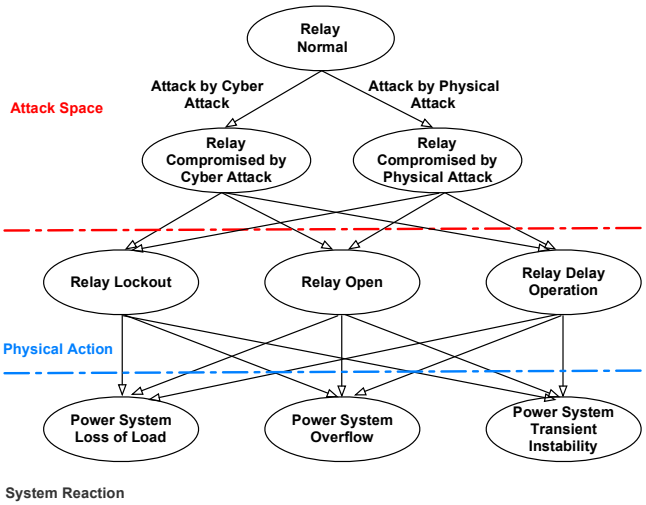


Fig. 1: The fused cyber-physical model in power systems regarding to digital protection relay

of five components. They are: States (S), Action (A), State Transition Model $P(s_{t+1}|s_t, a_t)$ that describes the transition of the environment state changes when the agent performs an action a in a current state s, Reward model $R(s_{t+1}|s_t, a_t)$, which describes the actual reward value that the agent receives from the environment after execution is performed, and the discount factor γ that controls the future rewards.

The value function $V(s)$ represents how beneficial it is for the agent to be in the state s. It is the expected total reward for an agent starting from state s, which depends on the policy π by which the agent picks actions For convenience, another function, the state-action pair function called the Q-function is also considered. The optimal Q-function $Q^*(s, a)$ means the expected total reward received by an agent starting in s and selecting action a. Hence, $Q^*(s, a)$ is an indication for how good it is for an agent to pick action a while in state s. It is equal to the summation of immediate reward after performing action a while in state s and the discounted expected future reward after the transition to a next state s'.

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V^*(s') \quad (1)$$

B. MDP Solver: Value and Policy Iteration

In value-iteration algorithm [22], the value function keeps improving until it converges. The objective of our problem is to find an optimal policy, and there is a chance that the optimal policy will converge before the value function. Therefore, policy-iteration algorithm [22] is also employed which improves the policy at each step and computes the value function according to this new policy until the policy converges. The algorithm for the MDP creation and solution is shown in Alg. 1. Value iteration is simpler but computationally heavy, but policy iteration is complicated but computationally cheap [22].

IV. Model for Optimal Response

A. Cyber-Physical Threat and Fault Model

Representation of cyber threats and physical faults is often an integral component for modeling optimal response based on

their impacts and risks involved [12]. In this paper, we model the threat by a sequential attack that adversely impacts the physical power system causing line overflow. The adversary is assumed to penetrate the Operational Technology (OT) network through a host(say $H2$) or the control network by initiating their targets onto a host in the Information Technology (IT) network (say $H1$) as shown in Fig. 3. Usually, there exist a few computers in the IT network, such as web servers in the Demilitarized Zone (DMZ) with lack of security patch upgrades, that could be hacked by attackers through simple buffer overflows or cross site scripting attacks. The attacker further intrudes inside control network by either compromising the firewall interfacing IT-OT or exploiting a vulnerable host, like the network printer spool service exploit in the Stuxnet attack, to modify the *Step7* software followed by centrifuge's PLC logic in a nuclear facility. With this penetration, they compromise the relays on the OT network to misoperate the breakers. Our solution assumes that the main control center where the EMS operates our decision making engine is considered trusted or uncompromised from cyber attacks.

For the fault model in physical networks, a contingency causing a short circuit in the transmission line that induces an abnormal current can trigger an overcurrent relay to open the circuit breaker. The generators output and the online load can affect the frequency and voltage stability in power systems. Any disturbances on those could trigger under frequency relay or under voltage relay to disconnect important components, like transformers or generators, to cause outages.

B. MDP Components

1) State Space: For the state space, we consider three types of states; they are purely cyber, purely physical, and mixed as shown in Fig. 2. Purely cyber states represent the compromise of the cyber components such as vulnerability exploits in cyber hosts. They are shown in orange in Fig. 2, representing states where cyber components are compromised, with the physical side not yet influenced. Purely physical states represent compromise of physical devices like mis-operation of circuit breaker (top layer of the Fig. 2). These states can be further classified into Normal, Alert and Emergency states [13]; here we consider only normal and emergency states based on line overflow. Intermediary states, like loss of load and generation which may not result in line overflow, are not considered. Mixed states refer to those where the devices which interface between the cyber and the physical side like relays, current transformers, etc. (bottom layer of Fig. 2). State C_{No} , P_{No} represent the normal $goal\_state$ where there is no cyber or physical compromise. In this state, the system can function without causing any disturbance or instability of the power supply in the system, and the power flow has already been re-dispatched from the previous compromised state.

2) Action Space: The action space can be classified into cyber and physical actions. The actions included to fix the firewall rules, patching a vulnerability or isolating a host from the network are a collection of possible cyber response actions. The physical response actions are either controlling the breakers or changing the time setting for the relays or controlling generation and load to reduce flow overflow. The possible actions in a given state are state dependent. For our case studies, we only consider 5 actions and they are $N\,oAction$, $P\,atchV\,uln$, $F\,W\,change$, $SwManual$ and $P\,hyAction$. Actions like $SwManual$ is to isolate the compromised cyber network immediately to avoid further influences and allow

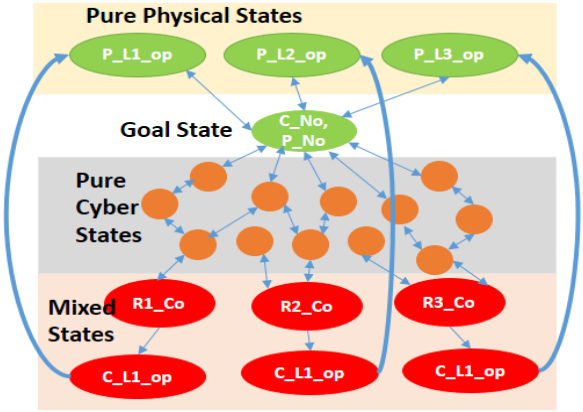


Fig. 2: State spaces of the MDP model

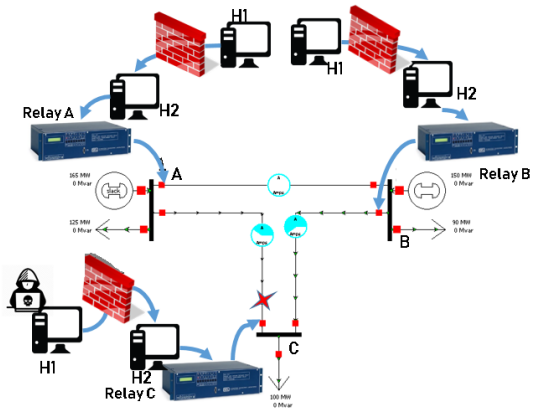


Fig. 3: Cyber-physical model for three substation case

operators to manually control physical devices to ensure the security and validity of the operation. One of the $P\,hyAction$ is to close the breaker manually. In our future work, we will delve into the granularity of each action in both cyber and physical domains. Deep granularity in determining the amount of set point for the generator or load will be determined by using an optimization solver.

3) Reward Model (R): The reward for different actions will be state dependent. For example, the reward for fixing a vulnerability and changing a firewall rules will differ. Vulnerabilities with high Common Vulnerability Scoring System (CVSS) scores have a severe impact [13], hence to pick actions that patch vulnerability with higher CVSS score is more rewarding. A change in firewall rule can impact the ongoing network traffic, hence its impact needs consideration. Similarly in power system, closing a breaker takes less time than adjusting generation or load to control line overflow. Our reward function depends on multiple criteria such as response time, cost and impact hence a simple weighted sum method as used in [18] is implemented. In [23], different forms of continuous and binary reward function used for the RL problem is discussed. Since the primary objective of our model is to protect physical components, more weight is given to the reward obtained from physical actions. For example, in our case we considered actions to prevent line overflow by controlling breakers. An action that results in higher flow deviation, is less rewarding. Since the type of reward function

can impact the optimal response, two different rewards are computed (2) using PowerWorld.

$$R1 = 1/\sum_{1}^{X} \Delta F^2 \quad or \quad R2 = 1/\sum_{1}^{X} |\Delta F|. \tag{2}$$

where ΔF is the deviation of flow from the normal operation.

4) State Transition Model (P): When an agent takes an action, the transition to a new state is dependent on the current state and also the type of the state. Some transitions may be purely deterministic and some may be stochastic(3).

$$P(s_{t+1}|s_t, a_t) = \begin{matrix} 0 \text{ or } 1 & s_t \in Ph_{states} \\ [0, 1) & s_t \in Cy_{states}, Mi_{states} \end{matrix} \tag{3}$$

In SOCCA [3], a static uncertainty based on CVSS scores [24] is used to determine P. Here we explore dynamic uncertainty, where the stochasticity is due to the existence of false positives and negatives in the alerts from Intrusion Detection Systems (IDS). We model the stochasticity using Dempster Shafer (D-S) Theory of uncertainty. The Dempster rule of combination [25] is used to compute the transition probability by combining evidences from multiple IDSes in a substation control network.

$$m_{1,2}(a) = \frac{1}{1-K} \sum_{b \cap c = a \neq \emptyset}^{X} m_1(b) * m_2(c) \tag{4}$$

where

$$K = \sum_{b \cap c = \emptyset}^{X} m_1(b) * m_2(c) \tag{5}$$

where, $m_1(b)$ is the mass function associated with the IDS1 for the hypothesis b (i.e. malicious packets directed to host H2) and $m_2(c)$ is the mass function associated with the IDS2 for the hypothesis c (i.e. H2 receive malicious packets from H1). Then $m_{1,2}(A)$ is the transition probability from H1 compromised state to both H1 and H2 compromised state at substation C in Fig. 3. The advantage of this theory over other Bayesian approaches are its ability to deal with the lack of prior probabilities for various events and the ability to combine evidences from multiple sources [26].

5) Discount Factor γ: This component affects how much importance the MDP solver gives to future rewards in the value function and ensures the rewards are bounded. A discount factor $\gamma = 0$ will result in state/action values with immediate reward. A higher γ represents the cumulative discounted future reward an agent expects to receive. Depending on the type of task, it will effect the convergence rate. In case of a continuous task γ must be between $[0, 1)$ and between $[0, 1]$ in discrete or episodic task.

V. Result and Analysis

A. Implementation

Three, five, and nine-substation cases were created using PowerWorld, and those were considered in modeling our optimal response problem. In each substation, we considered one bus and one relay that controls circuit breakers in the transmission lines connected to that bus. Each substation network is modeled to have a few cyber hosts. The network vulnerabilities in these hosts are exploited by attacker, based on the discovery of open ports using open-source network mapper NMAP scanner, to gain access into control network, finally to compromise the relay and mis-operate the breakers.

Algorithm 1 Pseudo code for MDP creation and solution

```
function BuildAndSolveMDP (PWcase, cyberInfo)
    Obtain branch L = L1, L2, L3 from PWcase
    Obtain normal flows F = F1, F2, F3 from PWcase
    Create States S = Cyb, Phy, Mix using PWcase and cyberInfo
    Define Actions A= NoAction, PatchVuln, FWchange, SwManual, PhyAction
    R = Compute the Reward for cyber states
    for each l in L do
        Open Breaker on l
        Solve Power Flow
        Compute reward using any one of Eq. 2
    end for
    P = Compute Transition Prob. for cyber states
    mdp = Tuple of < S, A, R, P, γ >
    π*_VI = Value_Iteration (mdp)
    π*_PI = Policy_Iteration (mdp)
    Run episodes using π*_VI and π*_PI for average rewards
end function
```

Fig. 4: Accuracy of optimal response for different cases with two types of rewards as mentioned in Eq. 2.

For reward modeling, the CVSS scores of these vulnerabilities were used from the National Vulnerability Database (NVD). Fig. 2 represents the state spaces of the MDP model created for this case. We test the optimal response problem, first exploring physical states and then considering all the types of states.

B. Considering Physical States

An initial model based on only physical states, is comprised of $L + 1$ states, where L is the number of transmission lines. The additional state is the normal state with no contingency. The action space comprised of $2L + 1$ states, where multiplication of two is due to the closed and open states of the breaker and the additional one is $NoAction$ where the agent takes no action. We tested the pure physical model for 3, 5 and 9 substation case and found that policy iteration found the optimal policies in fewer steps in comparison to value iteration method, as shown in Table I. It was observed that in the 9 substation case, single line contingencies did not cause line overflow, hence the reward function in Eq. 2 gave an inaccurate response as shown in Fig. 4.

C. Considering Physical, Cyber, and Cyber-physical States

Table II presents the number of iterations to obtain optimal policy using value and policy iteration when all the three types of states considered in the 3 and 5-substation model. Due to inaccuracy caused by the reward function for 9 substation case, we will test our engine for larger and complex system

TABLE I: Physical States: Value and Policy Iteration Comparison for 3,5, and 9 substation cases

Use Case	*States*	*Value Iteration*	*Policy Iteration*
3 subs	4	4	1
5 subs	6	5	1
9 subs	10	7	1

TABLE II: Physical, Cyber, and Cyber-physical States: Value and Policy Iteration comparison for 3 and 5 substation case

Use Case	*States*	*Value Iteration*	*Policy Iteration*
3 subs	17	74	2
5 subs	29	110	2

considering actions of generation and load control apart from breaker operation in future. The optimal policy obtained for this cyber-physical model is given in Table III. With corresponding current hybrid cyber-physical states, it provides the diagnose of the system and the optimal action to protect the system against adversaries.

After the optimal policy for different states is obtained, we run multiple episodes with starting state randomly picked with the $goal\_state$ fixed at C_No, P_No . The sequence of steps the agent would follow to reach the goal state, would determine the course of action the operator or the network administrator would take to address that contingent state. Further, we analyze the impact of different MDP components as discussed in Section IV.

D. Analysis of Discount Factor on Iteration Count

In our problem, we found that the discount factor affects the number of iterations the value iteration algorithm took to converge the value function for obtaining the optimal response. It was observed that with increasing discount factor, the number of iterations increases. Fig. 5(a) shows how the iteration count increased from an average value of 9 to 96 for change in γ from 0.1 to 0.98 for 5 substation case and from 9 to 77 for 3 substation case. Hence, we can say that the current optimal actions are determined not much on our future rewards. The policy iteration was converging in one or two iterations for varying γ.

E. Analysis of State Transition Model on Iteration Count

The State Transition Model impacts the number of iterations to reach an optimal policy. The self transition probabilities of the pure cyber states are affected by the transition probability to other states obtained using D-S Theory. Higher self transition probabilities would take more iteration for the value function to converge. From Fig. 5(b) shows the average number of value iterations increased from 52 to 104 for 3 substation case and from 68 to 115 for 5 substation case. This results convey that higher state transition between cyber states reduces the convergence time and number of iterations.

TABLE III: Computed Optimal Policy for a few states in 3 substation case

Type	Current State	Desc.	Opt. Action
P	C_No, P_No	No compromise	*NoAction*
C	C_H2BExp, P_No	H2 in Sub B expt	*PatchVuln*
C	C_H1CExp, P_No	H1 in Sub C expt	*PatchVuln*
C	$C\_H1B\ H2B\ Exp$	H1,H2 in Sub B expt	*PatchVuln*
CP	C_RAExp, P_Co	Relay in Sub A comp	*FWchange*
CP	$C\_Co, L1\_Op$	cyber attack L1 open	*SwManual*
P	$P\_Co, L1\_Op$	phy. fault L1 open	*PhyAction*

F. Impact of Reward Function Model on Iteration Count

The reward function model for the physical side is determined based on their impact on flows on the transmission lines. For 5 substation case, $R1$ from Eq. 2 fared better than $R2$ in terms of convergence and accuracy. For our simulation we gave higher weights to rewards for fixing a vulnerability than changing firewall rules, because a change in firewall rule may alter the network topology. We consider higher reward for physical actions rather than cyber actions since higher priority is to resolve line overflow then to address the cyber intrusion.

G. Analysis of State Transitions from Optimal Policy

Suppose at a given time, the agent is in state $C\_Co, L1\_Op$, then the optimal action would be to switch to operate the relay manually, then it reaches the next state $P\_Co, L1\_Op$ (based on the maximum reward from the R of the MDP, the agent get by taking the optimal action). From this state, the optimal action recommended is to close the breaker in transmission line $L1$ to reach the goal state C_No, P_No . Suppose the agent is in C_H2BExp, P_No state where the vulnerability of host H2 in Substation B is exploited and compromised. Optimal action recommended is to patch the vulnerability. Here there can be more actions such as removing the host H2 from the system but it may disrupt the communication.

H. Evaluation of Average Reward for varying Discount Factor

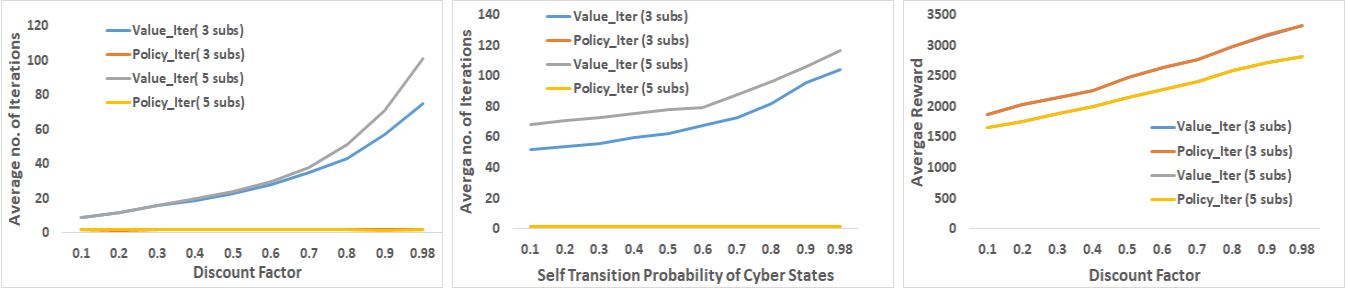
Once the optimal policy is determined, we run multiple episodes with a random selection of the starting states and see in how many steps it reaches the goal state which is C_No, P_No in our case. While modeling a complex system, we can have multiple intermediary goal states. To find the total average reward obtained by the agent by running an episode (6) is computed,

$$AverageReward = \frac{1}{E} \sum_{e=1}^{E} \sum_{t=1}^{S} \gamma^t * r(t) \qquad (6)$$

where $r(t)$ is the immediate reward, γ is the discount factor, E is the number of episodes and S is the number of steps taken by the agent to reach the goal state. We can observe from Fig. 5(c) the reward increasing with rise of discount factor.

VI. Conclusion

In this paper, we presented SCORE, a security-oriented cyber-physical optimal response engine that computes optimal action for different cyber-physical contingent state. It enables both operators and network administrators to take faster actions for system recovery. Our experiments on the three and five substation cases, helped us analyse the impact of different MDP components on the iteration count to validate the solution. We observed that a lower discount factor can help the agent get optimal policy in least number of value iteration. State transition model with lower self-transition probability, converged the value iteration method in fewer steps. With incorporation of larger case, SCORE faces state explosion issue, which we will address in future using Hierarchical MDPs. The convergence in case of Policy iteration is faster in comparison to Value iteration, because we have a limited set of policies in a given state for small power system cases. Moreover, with improvements of reward and state transition



(a) (b) (c)

Fig. 5: (a) Impact of discount factor on iteration count; (b) Impact of self transition probability on iteration count; (c) Impact of discount factor on average reward

model to support larger case, we envision SCORE to be an integral part of the next generation EMS system.

ACKNOWLEDGMENT

This research is supported by the US Department of Energy Cybersecurity for Energy Delivery Systems program under award DE-OE0000895 and a Grainger Foundation Frontiers of Engineering Grant.

REFERENCES

[1] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389–1407, 2017.

[2] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, March 2019.

[3] S. Zonouz, C. M. Davis, K. R. Davis, R. Berthier, R. B. Bobba, and W. H. Sanders, "Socca: A security-oriented cyber-physical contingency analysis in power infrastructures," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 3–13, Jan 2014.

[4] I. Ahamed, P. Nagendra Rao, and P. Sastry, "A reinforcement learning approach to automatic generation control," *Electric Power Systems Research*, vol. 63, pp. 9–26, 08 2002.

[5] M. Noroozian, G. Andersson, and K. Tomsovic, "Robust, near time-optimal control of power system oscillations with fuzzy logic," *IEEE Transactions on Power Delivery*, Jan 1996.

[6] A. Hoballah and I. Erlich, "Transient stability assessment using ann considering power system topology changes," in *2009 15th International Conference on Intelligent System Applications to Power Systems*, Nov 2009, pp. 1–6.

[7] I. Robandi, K. Nishimori, R. Nishimura, and N. Ishihara, "Optimal feedback control design using genetic algorithm in multimachine power system," *International Journal of Electrical Power Energy Systems*, vol. 23, no. 4, pp. 263 – 271, 2001.

[8] J. Soares, T. Sousa, Z. A. Vale, H. Morais, and P. Faria, "Ant colony search algorithm for the optimal power flow problem," in *2011 IEEE Power and Energy Society General Meeting*, 2011.

[9] C. G. Marcelino, P. E. Almeida, E. F. Wanner, M. Baumann, M. Weil, L. M. Carvalho, and V. Miranda, "Solving security constrained optimal power flow problems: a hybrid evolutionary approach," *Applied Intelligence*, vol. 48, no. 10, pp. 3672–3690, 2018.

[10] H. Mori and K. Takeda, "Parallel simulated annealing for power system decomposition," in *Conference Proceedings Power Industry Computer Application Conference*, May 1993.

[11] A. Sahu, H. N. R. K. Tippanaboyana, L. Hefton, and A. Goulart, "Detection of rogue nodes in ami networks," in *2017 19th International Conference on Intelligent System Application to Power Systems (ISAP)*, Sep. 2017, pp. 1–6.

[12] K. R. Davis, C. M. Davis, S. A. Zonouz, R. B. Bobba, R. Berthier, L. Garcia, and P. W. Sauer, "A cyber-physical modeling and assessment framework for power grid infrastructures," *IEEE Transactions on Smart Grid*, Sep. 2015.

[13] K. R. Davis, R. Berthier, S. Zonouz, G. Weaver, R. B. Bobba, E. Rogers, P. W. Sauer, and D. M. Nicol, "Cyber-physical security assessment for electric power systems," in *IEEE-HKN: The Bridge*, 2017.

[14] "North American Electric Reliability Corporation - Reliability Concepts," https://www.nerc.com/files/concepts_v1.0.2.pdf.

[15] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2444–2453, 2015.

[16] J. Hong, R. Nuqui, D. Ishchenko, Z. Wang, T. Cui, A. Kondabathini, D. Coats, and S. Kunsman, "Cyber-physical security test bed: A platform for enabling collaborative cyber defense methods," in *PACWorld Americas Conference*, 2015.

[17] H. Huang and K. Davis, "Power system equipment cyber-physical risk assessment based on architecture and critical clearing time," in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2018, pp. 1–6.

[18] S. Iannucci and S. Abdelwahed, "A probabilistic approach to autonomic security management," in *2016 IEEE International Conference on Autonomic Computing (ICAC)*, July 2016.

[19] S. Ossenbhl, J. Steinberger, and H. Baier, "Towards automated incident handling: How to select an appropriate response against a network-based attack?" in *2015 Ninth International Conference on IT Security Incident Management IT Forensics*.

[20] J. Yan, H. He, X. Zhong, and Y. Tang, "Q-learning-based vulnerability analysis of smart grid against sequential topology attacks," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 200–210, Jan 2017.

[21] S. A. Zonouz, H. Khurana, W. H. Sanders, and T. M. Yardley, "Rre: A game-theoretic intrusion response and recovery engine," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 2, Feb. 2014.

[22] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[23] L. Matignon, G. Laurent, and N. Fort-Piat, "Reward function and initial values: Better choices for accelerated goal-directed reinforcement learning," in *Artificial Neural Networks 2006*.

[24] J. Li, X. Ou, and R. Rajagopalan, *Uncertainty and Risk Management in Cyber Situational Awareness*, 2010, vol. 46.

[25] L. A. Zadeh, "A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination," *AI Mag.*, Jul. 1986.

[26] L. Zomlot, S. C. Sundaramurthy, K. Luo, X. Ou, and S. R. Rajagopalan, "Prioritizing intrusion analysis using dempster-shafer theory," in *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*, ser. AISec '11. ACM, 2011.