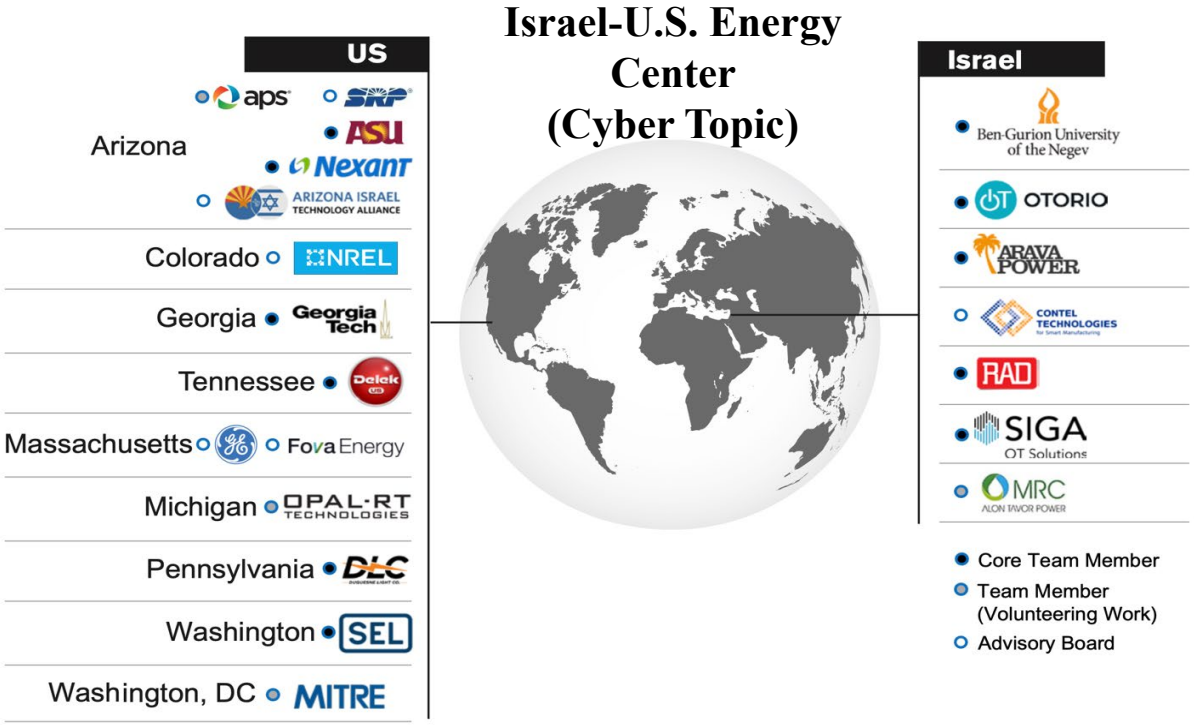
# Comprehensive Cybersecurity Technology for Critical Power Infrastructure AI-Based Centralized Defense and Edge Resilience





Prepared for

**Itai Ganzer** and **Ofer Goldhirsh**

Israel Innovation Authority

**Avi Shavit** and **Eynan Lichterman**

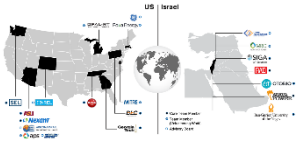
Israel Ministry of Energy


## Task 16: Reinforcement Learning Control for Cyber Physical Systems


Mohammadamin Moradi
on behalf of
Ying-Cheng Lai
Arizona State University

8/25/2022

# Preference Based Resource Allocation in CPS Using DRL



**Preferences**:

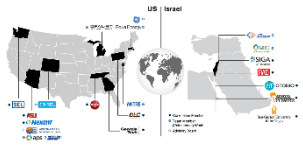
- Preferences and priorities play a key role in the real-world decision-making problems

**Examples:**

- In Chess, an AI player wants to win. It also prefers to win without losing its **queen**. Moreover, it prefers losing the **rooks** rather than losing the **knights**.

- In cybersecurity of power grids, the government wants to keep transmission lines safe; however, if a blackout happens due to attacks, it prefers blackouts in less populated/significant areas.

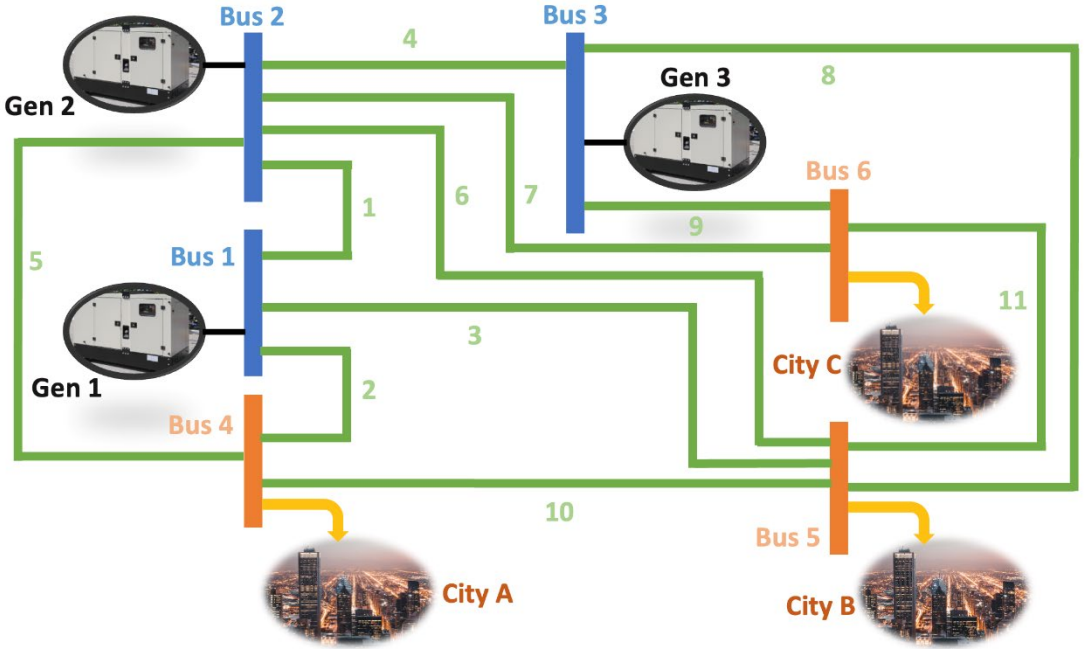
# Preferences in a Power Grid



- Attacker attacks generators. Defender defends them.

**Attacker's Preferences ($P$):**

- $PA$: The more generators being attacked the better.
- $PB$: If attacking all generators is possible, attacking Gen 3 in the end is preferred.
- $PC$: Attacking Gen 1 first is preferred to attacking other generators first.




**W&W 6-Bus System**

From **defender's Point of View**, preferences can be the complement of the above ($P°$).

- $PA°$: The less generators being attacked the better.
- $PB°$: If defending all generators is not possible, Gen 3 going down first or second is preferred.
- $PC°$: Gen 1 not going down first is preferred to other generators going down first.

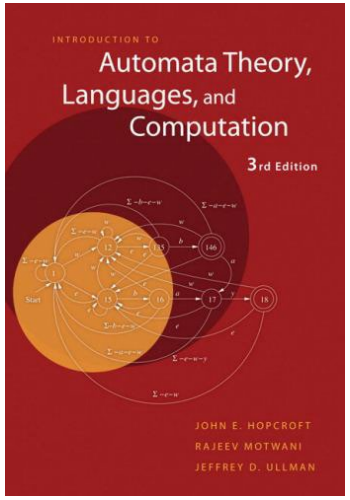
# Preference Transformed into Mathematical Formula

**Attacker's Preferences ($P$):**

- $PA$: The more generators being attacked the better.
- $PB$: If attacking all generators is possible, attacking Gen 3 in the end is preferred.
- $PC$: Attacking Gen 1 first is preferred to attacking other generators first.

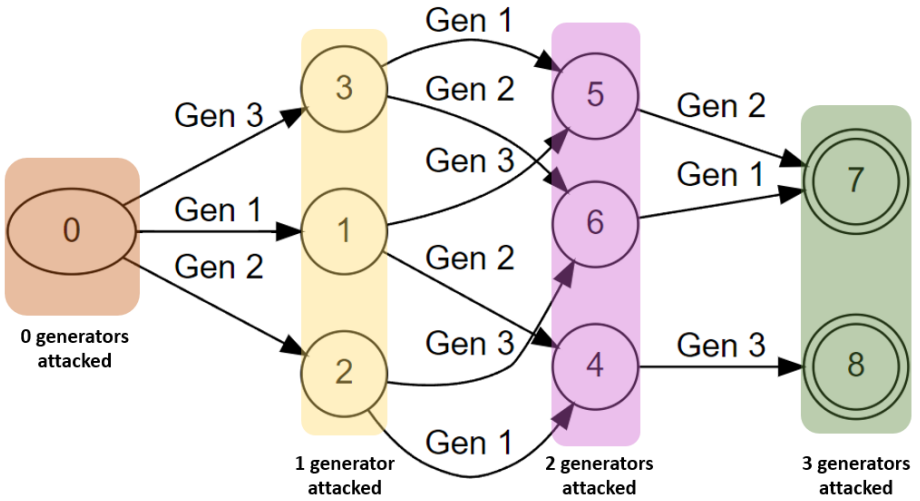
## Using Automata Theory:

- $\phi_1 := \{8\} \geq \{7\} \geq \{4,5,6\} \geq \{1,2,3\} \geq \{0\}$
- $\phi_2 := \{1\} \geq \{2,3\}$




- **Value of Preference Satisfaction:**
- $V_{PS}$ is closely related to the probability of occurring of that preference.
- $V_{PS}$ for a preference formula $X_0 \leq X_1 \ldots \leq X_n$ is defined as $P(X_i)$ if there exists some $i$ such that $P(X_i) \geq P(X_{i-1})$ while for all $k \geq i$ the $P(X_k) < P(X_{k-1})$ holds and is zero if otherwise.

- Policy $\pi_1$ satisfies the preference by %50 while policy $\pi_2$ satisfies the preference by %80

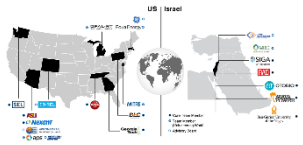



**Deterministic Finite Automaton**

| Automaton state set | $P_{\pi_1}$ | $P_{\pi_2}$ |
|---|---|---|
| $\{8\}$ | 0.05 | 0.8 |
| $\{7\}$ | 0.15 | 0.05 |
| $\{4,5,6\}$ | 0.5 | 0 |
| $\{1,2,3\}$ | 0.1 | 0.05 |
| $\{0\}$ | 0.2 | 0.1 |

$\phi_1 := \{8\} \geq \{7\} \geq \{4,5,6\} \geq \{1,2,3\} \geq \{0\}$

# Optimization Problem Formulation



- **Problem:**
**Satisfy the preferences as much as possible**

- Formulate the problem as a Mixed Integer Program (MIP)

Constraints are constructed from the

- definition of $V_{PS}$,
- the MDP equations(power grid simulated through DCSIMSEP) and the
- automaton made from preferences

**MIP problem:** $\max_{B,y,V_{PS}} V_{PS}$

$$subject\ to\ Eqs.\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8$$

**Constraints:**

$$0 \leq V_{PS} \leq B \quad (2)$$

$$B - 1 \leq V_{PS} - y(T, P') \leq 0 \quad (3)$$

$$B(1+\epsilon) + 1 \leq y(T, P') - y(T, P) \leq B(1+\epsilon) - \epsilon \quad (4)$$

$$B\ is\ a\ binary \quad (5)$$

$$all\ y\ are\ non - negative, \quad (6)$$

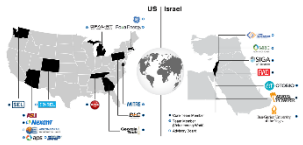
$$\sum_{a \in A} y(0, (\tilde{s}, s), a) = d(\tilde{s}, s) \quad (7)$$

$$\sum_{a \in A} y(t, (\tilde{s}', s'), a) = \sum_{a \in A} \sum_{(\tilde{s}, s) \in \tilde{S} \times S} \Delta((\tilde{s}', s') | (\tilde{s}, s), a) y(t-1, (\tilde{s}, s), a). \quad (8)$$

$$\min_{B,y,V_{PS}} -V_{PS}$$

$$s.t.: \ y1 + y2 = 0.9$$
$$y3 + y4 = 0.05$$
$$y5 + y6 = 0.05$$
$$y7 + y8 = 0$$
$$y9 + y10 - 0.05y1 - 0.05y2 - 0.05y6 - 0.05y3 = 0$$
$$y11 + y12 - 0.8y1 - 0.1y2 - 0.95y3 = 0$$
$$y13 + y14 - 0.1y1 - 0.8y2 - 0.95y6 = 0$$
$$y15 + y16 - 0.05y1 - 0.05y2 - y4 - y5 - y7 - y8 = 0$$
$$y17 + y18 - 0.05y9 - 0.05y10 - 0.05y14 - 0.05y11 = 0$$
$$y19 + y20 - 0.8y9 - 0.1y10 - 0.95y11 = 0$$
$$y21 + y22 - 0.1y9 - 0.8y10 - 0.95y14 = 0$$
$$y23 + y24 - 0.05y9 - 0.05y10 - y12 - y13 - y15 - y16 = 0$$
$$y25 + y26 - 0.05y17 - 0.05y18 - 0.05y22 - 0.05y19 = 0$$
$$y27 + y28 - 0.8y17 - 0.1y18 - 0.95y19 = 0$$
$$y29 + y30 - 0.1y17 - 0.8y18 - 0.95y22 = 0$$
$$y31 + y32 - 0.05y17 - 0.05y18 - y20 - y21 - y23 - y24 = 0$$
$$-V_{PS} + y27 + y28 + B \leq 1$$
$$V_{PS} - y27 - y28 \leq 0$$
$$y27 + y28 - y29 - y30 - (1+\epsilon)B \leq -\epsilon$$
$$-y27 - y28 + y29 + y30 + (1+\epsilon)B \leq 1$$
$$-V_{PS} \leq 0$$
$$V_{PS} - B \leq 0$$
$$-y_i \leq 0 \quad i = 1, \ldots, 32$$
$$B \in \{0, 1\}$$

# Resource Allocation



- Constraints are constructed from the definition of $V_{PS}$, the MDP equations(the power grid simulated through DCSIMSEP) and the automaton made from preferences

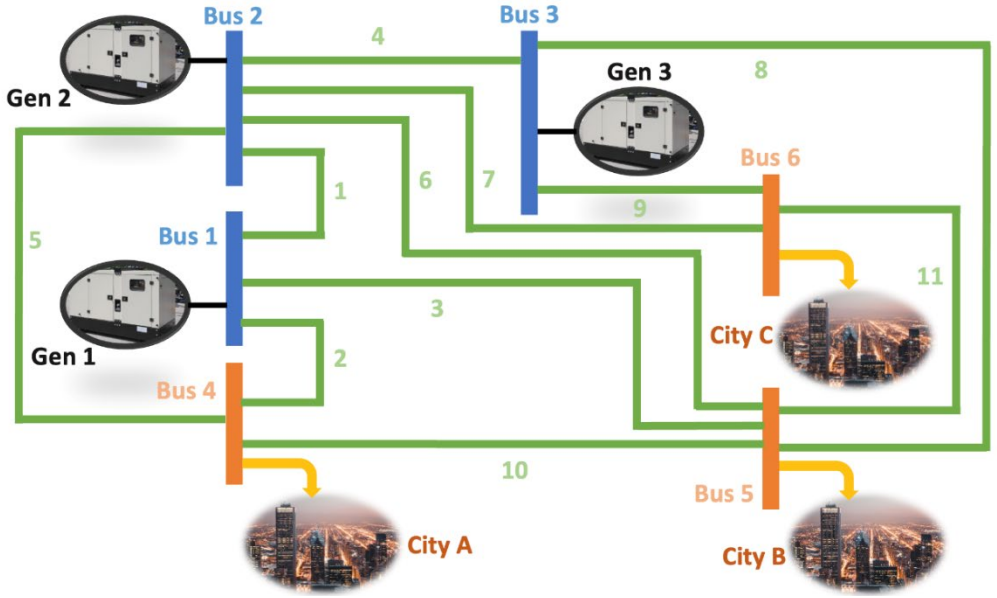
Resource Allocation:

- Defender has limited resources (funding, soldiers, etc.)
- The probability of attack success depends on allocated resources

$$p(i) = \frac{1}{1 + h(i)}$$

- Different resource allocation changes MDP equations

**New Resource Allocation** → **New MDP equations** → **New Constraints** → **New MIP**

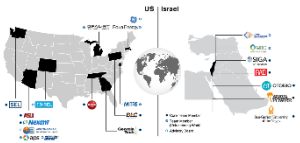



**MIP problem:**

$$\max_{B,y,V_{PS}} \; V_{PS}$$

$$subject\ to\ Eqs.\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8$$

# Deep RL in Preference Based Resource Allocation

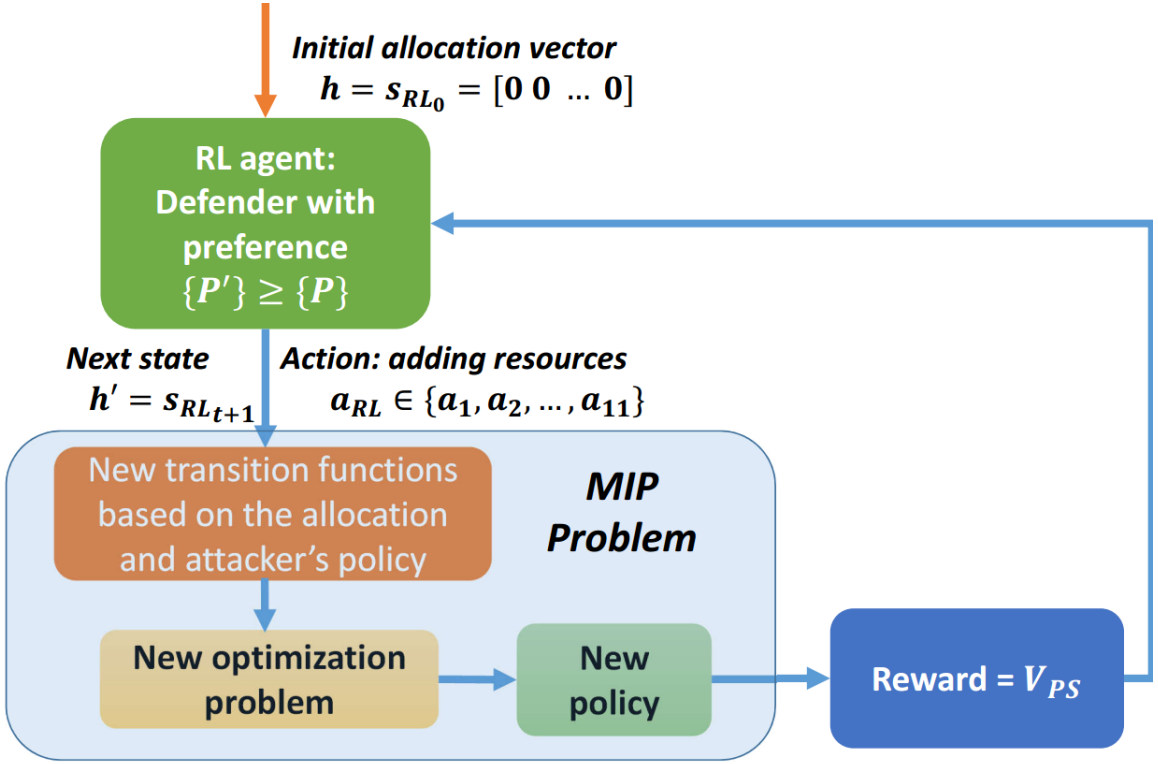


**New Resource Allocation** → **New MDP equations** → **New Constraints** → **New MIP**

- **Problem:**

**Optimally allocate resources to defend transmission lines such that the allocation satisfies the preferences as much as possible**

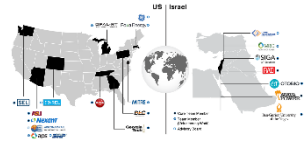
- Use Deep Reinforcement Learning (DRL) to learn the optimal resource allocation
- At each time step, a new MIP will be constructed and solved




**MIP problem:** $\max_{B,y,V_{PS}} V_{PS}$

$subject\ to\ Eqs.\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8$

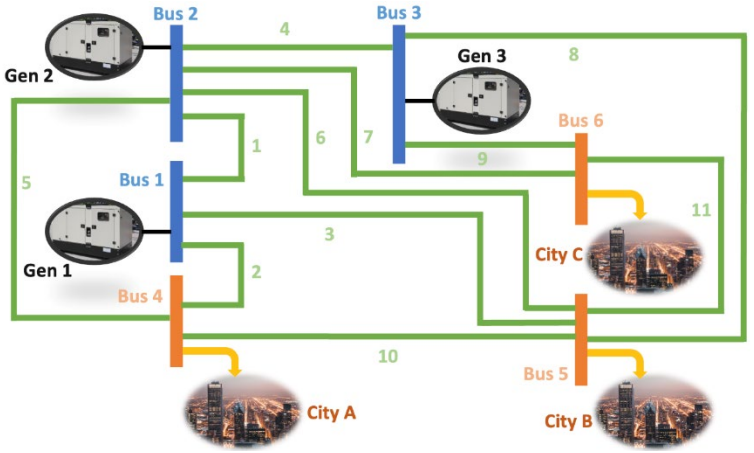
# Results



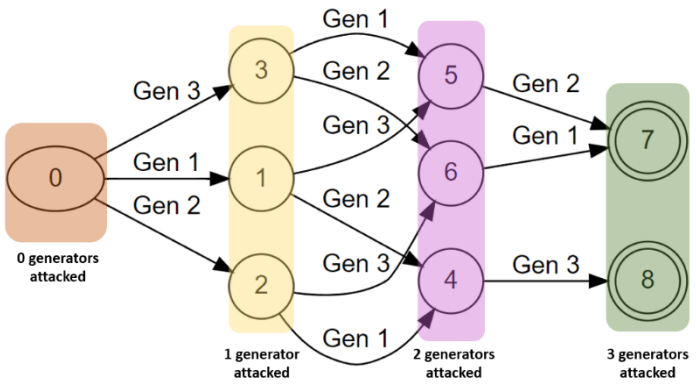
- **Problem:**

Optimally allocate resources to defend transmission lines and satisfy the preferences as much as possible

- **Power Grid: W&W 6-Bus System**
- **DRL method: DQN**
- **Preferences considered:**
- $\#1: \{6\} \leq \{0,1,2,3,4,5\}$
- $\#2: \{5,6,7,8\} \leq \{0,1,2,3,4,5\}$
- $\#3: \{4,5,6,7,8\} \leq \{0,1,2,3\}$

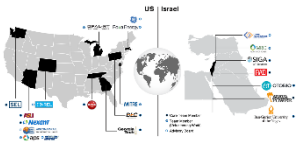



**Simulation tool**: DC load flow simulator of cascading (separation) in power systems (DCSIMSEP)




| Resources (H) | Pref.# | Allocation Vector (h) | $V_{PS}$ |
|---|---|---|---|
| H=0 | 1 | h=[0,0,0,0,0,0,0,0,0,0,0,0] | 0.3192 |
| H=0 | 2 | h=[0,0,0,0,0,0,0,0,0,0,0,0] | 0.3192 |
| H=0 | 3 | h=[0,0,0,0,0,0,0,0,0,0,0,0] | 0 |
| H=1 | 2 | h=[0,0,0,0,1,0,0,0,0,0,0,0] | 1 |
| H=2 | 2 | h=[0,0,0,0,1,0,0,0,0,0,1,0] | 1 |
| H=3 | 2 | h=[0,0,0,1,1,1,0,0,0,0,0,0] | 1 |
| H=4 | 1 | h=[0,0,0,2,1,1,0,0,0,0,0,0] | 1 |
| H=7 | 3 | h=[0,0,0,0,3,0,2,2,0,0,0,0] | 0.5941 |
| H=10 | 3 | h=[0,0,1,0,4,0,2,3,0,0,0,0] | 0.6057 |
| H=20 | 3 | h=[0,0,0,0,10,0,1,9,0,0,0,0] | 0.6475 |

# Products



## Defending smart electrical power grids against cyberattacks with deep Q-learning

Mohammadamin Moradi,[1] Yang Weng,[1] and Ying-Cheng Lai[1,2,*]

[1] *School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA*

[2] *Department of Physics, Arizona State University, Tempe, Arizona 85287, USA*

(Dated: July 19, 2022)

A key to ensuring the security of smart electrical power grids is to devise and deploy effective defense strategies against cyberattacks. To achieve this goal, an essential task is to simulate and understand the dynamical interplay between the attacker and defender, for which stochastic game theory and reinforcement learning stand out as a powerful mathematical/computational framework. Existing works were based on conventional Q-learning to find the critical sections of a power grid to choose an effective defense strategy, but the methodology was applicable to small systems only. Additional issues with Q-learning are the difficulty to consider the timings of cascading failures

Status: Submitted to PRX Energy

## Preference based resource allocation in cyber-security defense of power grids using reinforcement learning

Mohammadamin Moradi,[1] Yang Weng,[1] and Ying-Cheng Lai[1,2,*]

[1] *School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA*

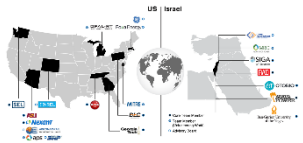
[2] *Department of Physics, Arizona State University, Tempe, Arizona 85287, USA*

(Dated: August 10, 2022)

**Abstract.** Preferences and priorities play a key role in the real world decision making problems. Moreover, limited human/financial resources in cyber-security applications highlight the significance of an optimal resource allocation. When combined, a preference based optimal resource allocation problem proves itself worthy in the decision making field. In this paper, we propose reinforcement learning based framework to solve the mentioned problem. Our solution, uses Automata theory

Status: Draft version ready

# Next



Basic Research:

- Larger power grids or more complex preferences cause the number of constraints to grow exponentially
- Large number of constraints renders the MIP to be infeasible by conventional solvers
- Find a way to deal with large number of constraints

Commercialization:

- Work with Nexant to implement the principle and methodologies of reinforcement learning control of cyber physical systems into the existing industrial Operational Technology and Industrial Control Systems management software tools.