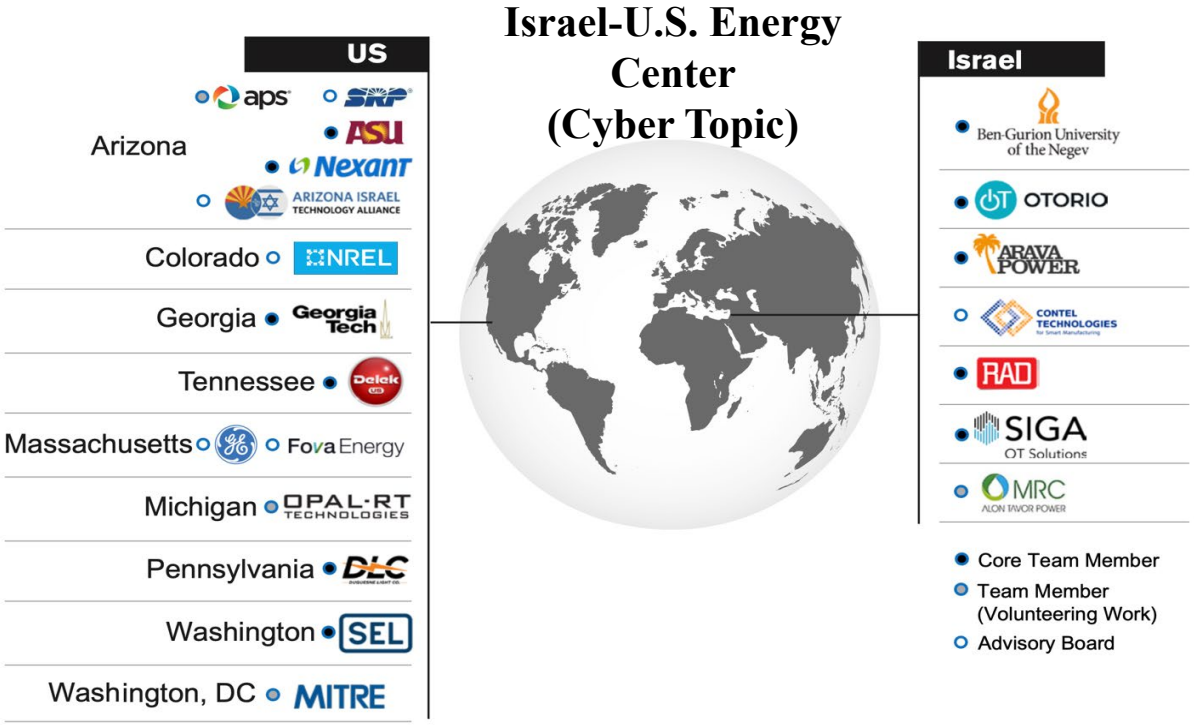
Comprehensive Cybersecurity Technology for Critical Power Infrastructure AI-Based Centralized Defense and Edge Resilience



Prepared for

Itai Ganzer and **Ofer Goldhirsh**

Israel Innovation Authority

Avi Shavit and **Eynan Lichterman**

Israel Ministry of Energy

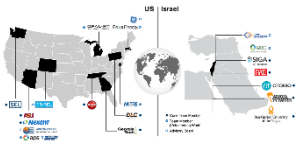
Task 16: Reinforcement Learning Control for Cyber Physical Systems

Ying-Cheng Lai

Arizona State University

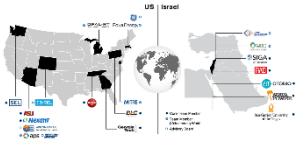
5/9/2022

Overview of Task Progress to Date



Task 16: Reinforcement Learning Control for Cyber-Physical Systems (CPS)					
PI: Ying-Cheng Lai					
Task Milestones	Start Date	End Date	Today	Total Duration in Days	Progress
M16.1 Demonstration of ML-based digital twin to simulate diverse control scenarios	5/1/2022	10/31/2022	4/27/2022	183	0%
M16.1 Actual Progress	1/5/2022	10/31/2022	4/27/2022	299	**37%**
M16.2 Develop reinforcement learning based criteria for selecting the "best" control scenarios	11/1/2022	4/30/2023	4/27/2022	180	0%
M16.2 Actual Progress	1/5/2022	4/30/2023	4/27/2022	480	**23%**
M16.3 Construct a control-scenario library to correspond with attack-scenario library	5/1/2023	10/30/2023	4/27/2022	182	0%
M16.3 Actual Progress	5/1/2023	10/30/2023	4/27/2022	182	0%
M16.4 Demonstrate the power of control-scenario library to generate quick response	11/1/2023	4/30/2024	4/27/2022	181	0%
M16.4 Actual Progress	11/1/2023	4/30/2024	4/27/2022	181	0%
M16.5 Implement control-scenario library in OT and ICS software	5/1/2024	10/30/2024	4/27/2022	182	0%
M16.5 Actual Progress	5/1/2024	10/30/2024	4/27/2022	182	0%

Technical Content: Machine Learning Guaranteed Security of Power Grids



- Scenario: Defense management team of a given large power grid performs stochastic game-playing to simulate the dynamic interplay between the attacker and the defender;
- Goal: to uncover the “best” attack strategies that can result in the maximal damage to the grid;
- Optimal defense (control) strategy: protecting the components in the grid that such attack strategies target;
- Mathematical model: treating attacker-defender interaction as a zero-sum game;
- Approach: Deep Q-learning with a customized reward function for achieving the desired objectives as directly as possible;
- Cascading failures: The deep-Q learning framework can be used to address problems of cascading failures and timing delays.

Technical Content: Q-Learning vs Deep Q-Learning

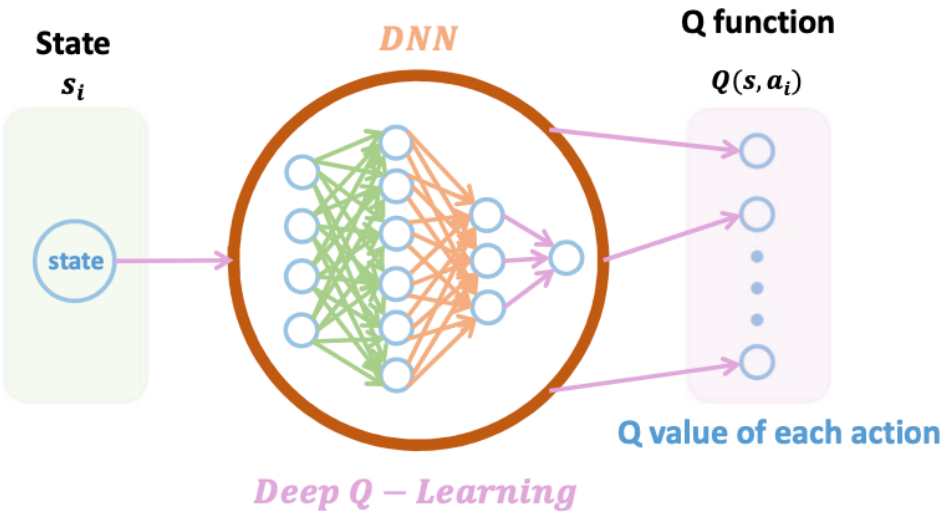
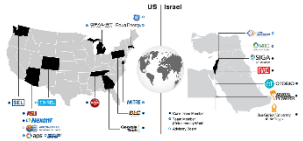


FIG. 1. Q-Learning versus deep Q-Learning. The implementation of the Q-table is the main difference between Q-learning and deep Q-learning. Instead of mapping a (state, action) pair to a Q-value using Q-table as is done in Q-learning, deep Q-learning uses neural networks to map the states to (action, Q-value) pairs, which is the core reason deep Q-learning can be used to solve large scale problems.

Technical Content: Attack Scenario and Simulation Method



Attack on smart power grids:

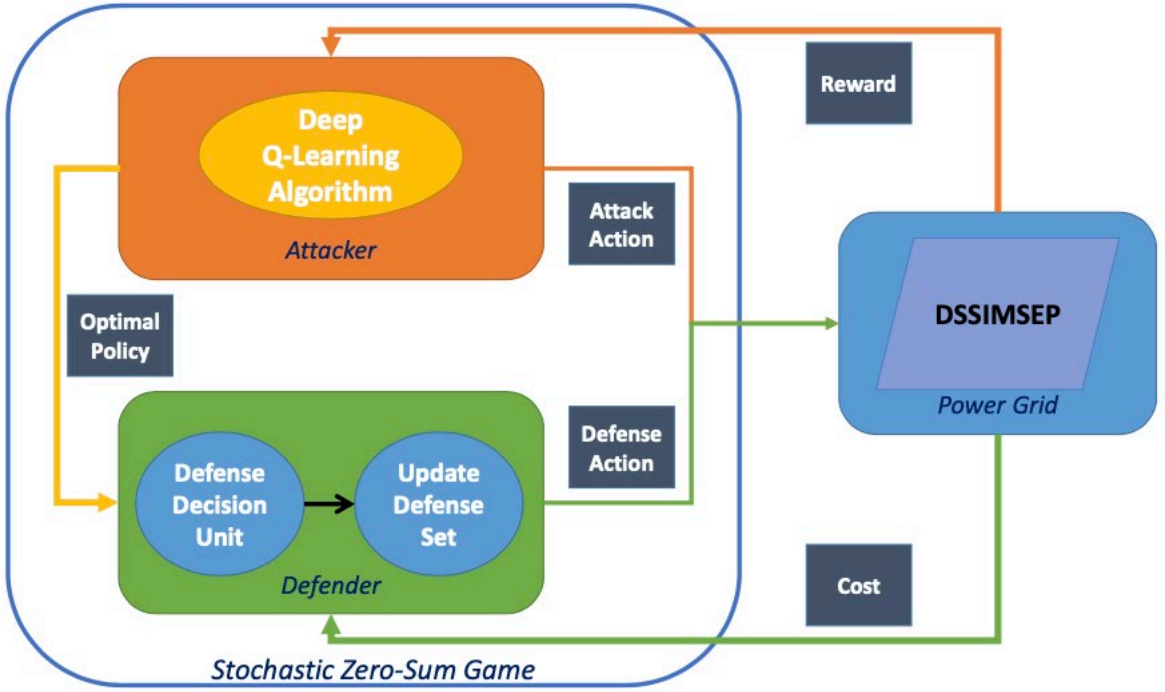
- Attacker attempts to cause a pre- determined percentage of the transmission lines to go outage
- Attacker attempts to maximize the generation loss in the power system through a sequence of attacks;

Defense: the defender strives to mitigate the attack consequences, regardless of whether they are due to transmission line outages or are caused by generation loss;

Generation loss $$G_{\text{loss}}^{-} = G_{\text{loss}}^{\text{init}} * \frac{t_{\text{cas}}}{T} + G_{\text{loss}}^{\text{std}} \frac{T - t_{\text{cas}}}{T}$$

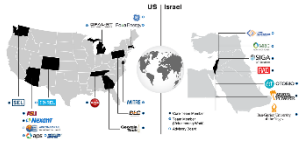
- T: time at which next attack will be launched
- G_{loss}^{init}: generation loss caused initially by the attack
- G_{loss}^{std}: generation loss during actions
- t_{cas}: cascading failure length caused by the attack

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



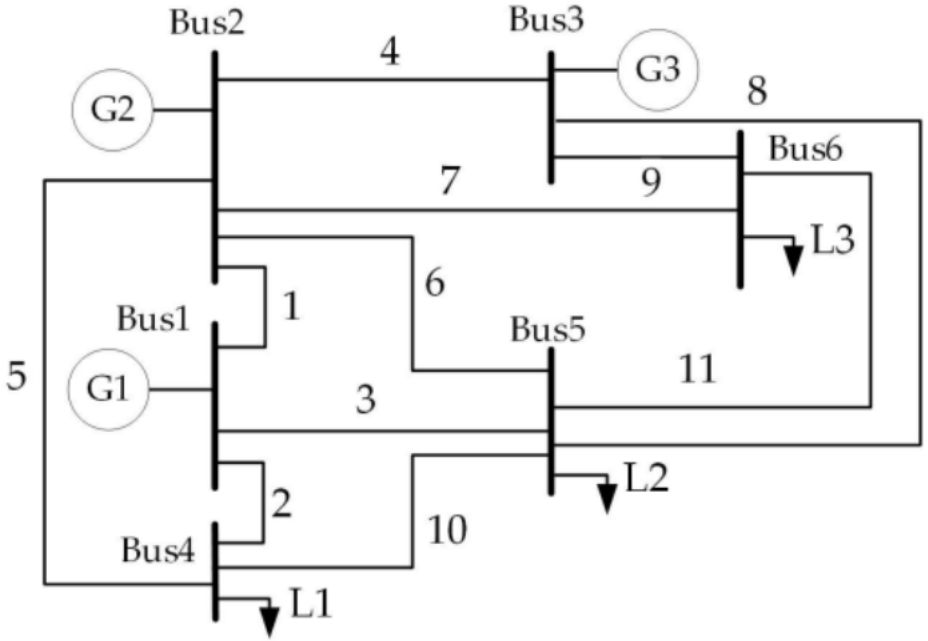
- Attacker uses the deep Q-learning algorithm to find an optimal attack sequence to maximize the generation loss/transmission line outage;
- Defender updates its defense set based on attacker's previous policy;
- The chosen actions of both players are given to the DCSIMSEP power flow simulator and reward/cost is calculated and returned to the players;
- The process continues until the defender's protection set remains unchanged for a number of cycles.

Technical Content: Attack on Smart Power Grids



The attacker attempts to cause a pre-determined percentage of the transmission lines to go outage

Wood and Wollenberg six-bus system



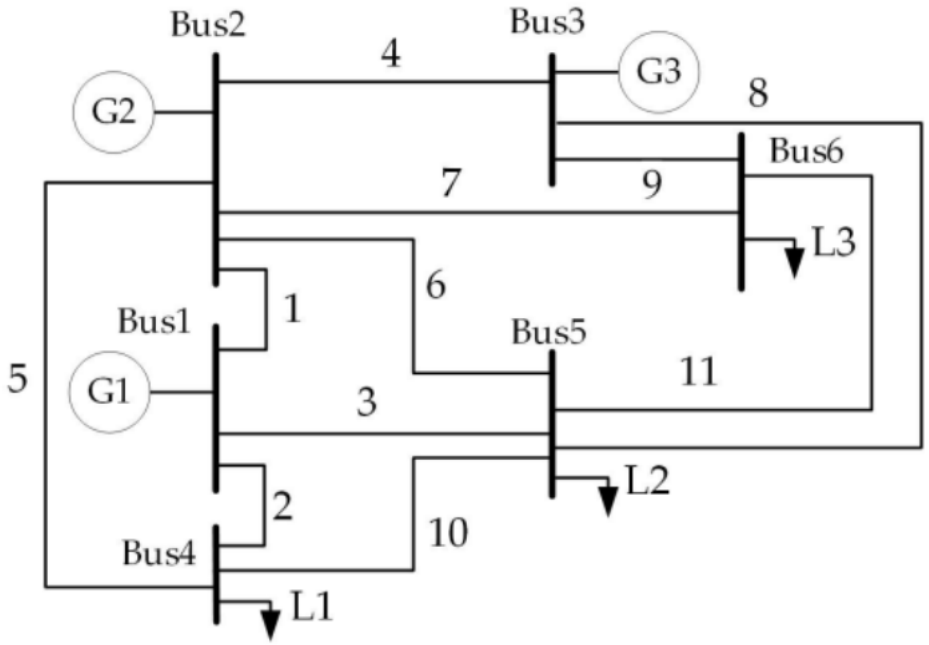
Reward function:

$$r = \begin{cases} r_1, & \text{for IO} > \text{AO} \\ r_2, & \text{if attack is final} \\ \text{IO/AO}, & \text{otherwise,} \end{cases} \qquad r_1 > r_2$$

- IO: the instant number of transmission line outages caused by the attack;
- AO: the attack objective
- Example: if the protection set consists of lines 1 and 2, attacking line 5 will cause an instant outage of five lines (IO = 5 > AO = 4):
  → reward = r_1
- Attacking line 3 will cause lines 1, 2, and 3 to go down
  → reward = $3/4$.
- Eventually, if the number of current downed transmission lines is less than AO and an attack causes the number of downed lines to be equal to or larger than AO
  → reward = r_2

Technical Content: Attack Induced Generation Loss

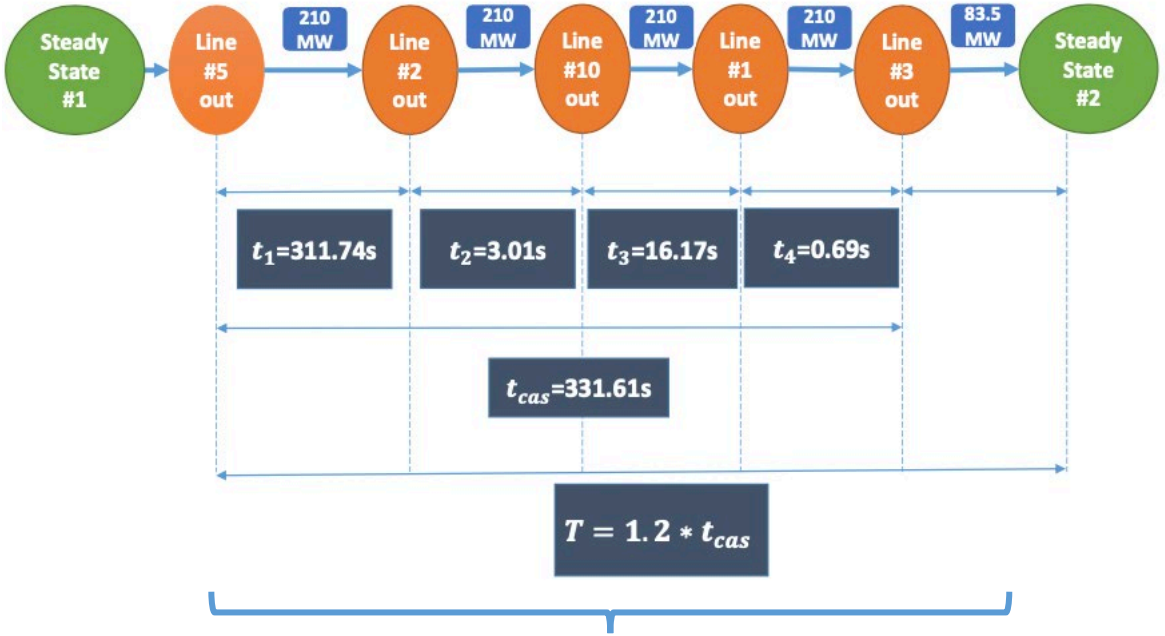
Wood and Wollenberg six-bus system



- If the protection set consists of lines 1 and 2, attacking line 5 will cause an instant outage of five lines (IO = 5 > AO = 4): reward = r_1,

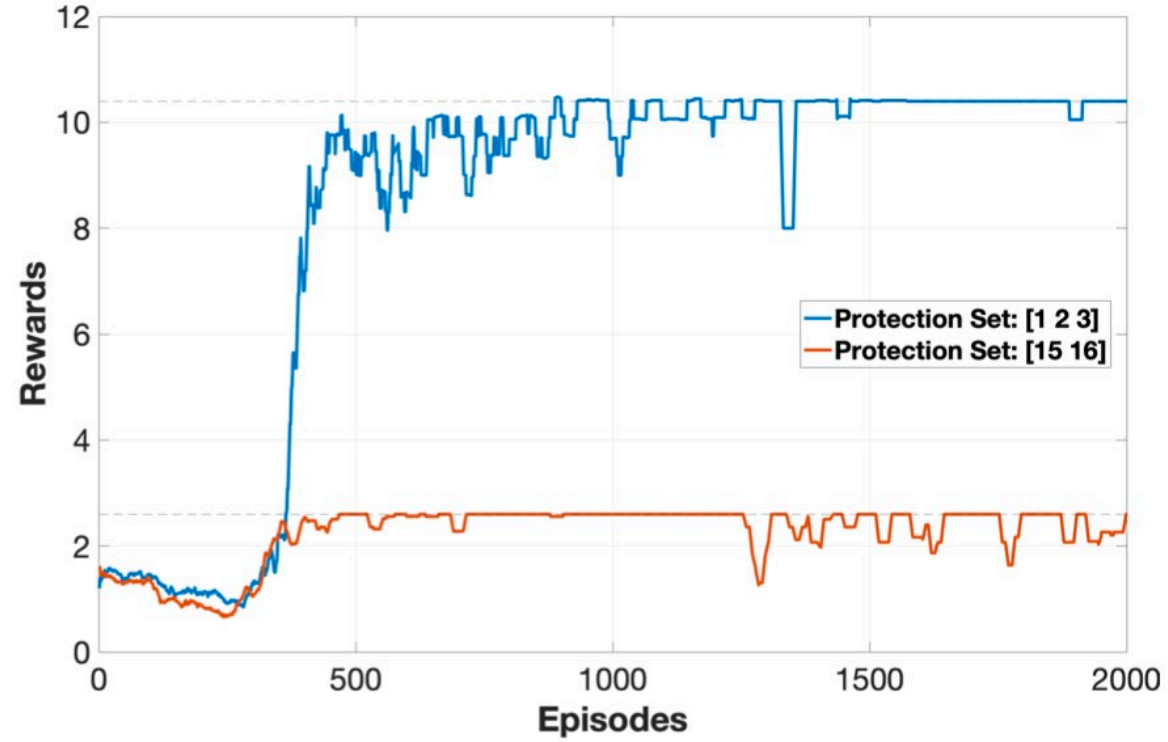
$$G_{loss}^{init} = 210 \text{ MW}, G_{loss}^{std} = 84 \text{ MW}, t_{cas} = 331.61 \text{ s}$$

$$G_{\text{loss}}^{-} = G_{\text{loss}}^{\text{init}} * \frac{t_{\text{cas}}}{T} + G_{\text{loss}}^{\text{std}} \frac{T - t_{\text{cas}}}{T} = 167.83 \text{ MW}$$



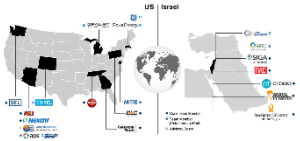
Transient dynamics

Technical Content: Optimal Defense Strategy for IEEE 30-Bus System



- Evolution of reward function values during the learning phase in the switching line problem
- When the defender chooses a random protection set {1, 2, 3}, the attacker finds an optimal sequence to get a large reward;
- After some cycles, the defender chooses {15, 16} as its protection set and the attacker is unable to find a sequence with a reward of more than r = 2.6.

Product: A Working Manuscript to be Submitted Shortly



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*

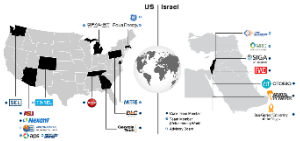
(Dated: May 3, 2022)

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 is applicable to small systems only. Additional issues with Q-learning are the difficulty to take into account the timings of cascading failures in the reward function and deterministic modeling of the game while the attack success depends on parameters and typically has a stochastic nature. Our solution to overcoming these difficulties is to develop a deep Q-learning based stochastic zero-sum Nash strategy solution. We demonstrate the workings of our deep-Q learning solution using the benchmark W&W 6-bus and the IEEE 30-bus systems, the latter being a relatively large scale power-grid system that defies the conventional Q-learning approach. Comparison with alternative reinforcement learning methods provides further support for the general applicability of our deep-Q learning framework in ensuring secure operation of modern power grid systems.

ACKNOWLEDGMENT

This work was mainly supported by U.S.-Israel Energy Center managed by the Israel-U.S. Binational Industrial Research and Development (BIRD) Foundation.

Commercialization Plan



Past encounter with commercialization (Lai):

- In August 2020, the Air Force/MIT Artificial Intelligence Accelerator launched a public challenge to help create the artificial intelligence needed to solve the magnetic navigation problem.
- The specific call is for the signal enhancement for magnetic navigation (MagNav) challenge problem with the goal to use magnetometer readings recorded from within a cockpit and remove the aircraft magnetic noise to yield a clean magnetic signal.
- The ASU team led by Lai responded and tested three types of machine learning methods: multilayer perceptrons (MLPs), reservoir computing, and long short-term memory (LSTM) neural networks.
- In December 2020, the Air Force/MIT Artificial Intelligence Accelerator placed the ASU team as the winning team.
- The involved Air Force officers suggested to Lai commercializing the machine-learning technique.

The ASU Task 16 team will 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.