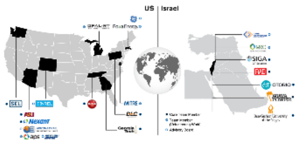
Tasks 5 and 8: OT Cyberattacks Identification and Mitigation

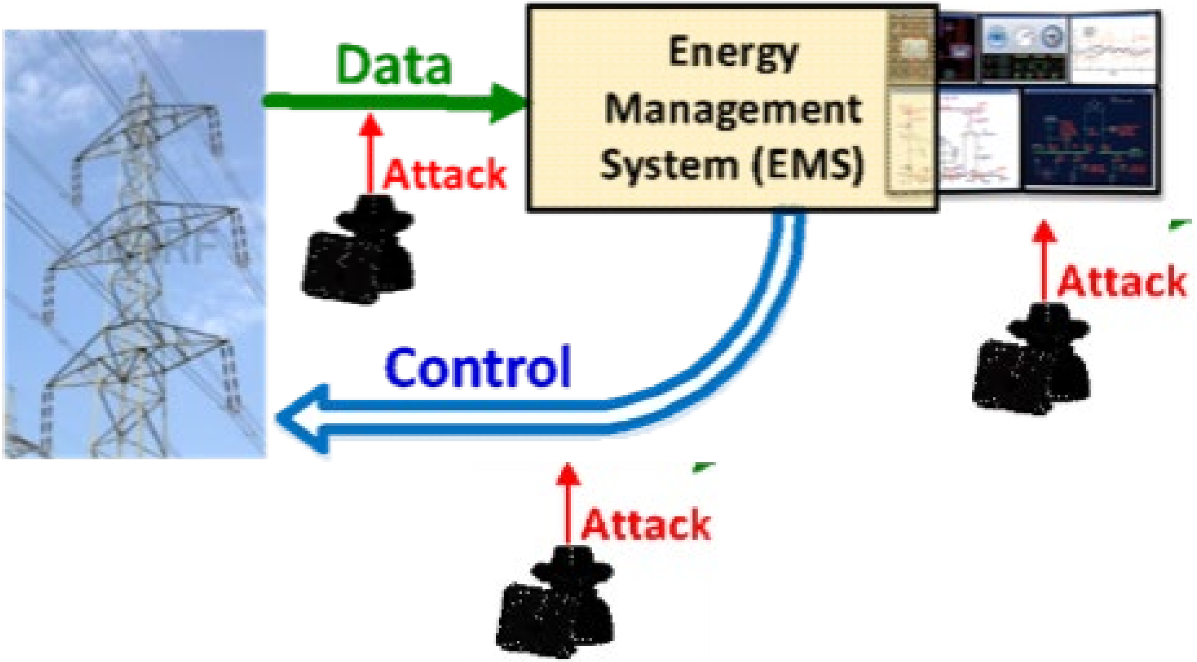
Lalitha Sankar

Arizona State University

1/24/2022

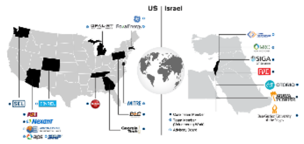
Motivation

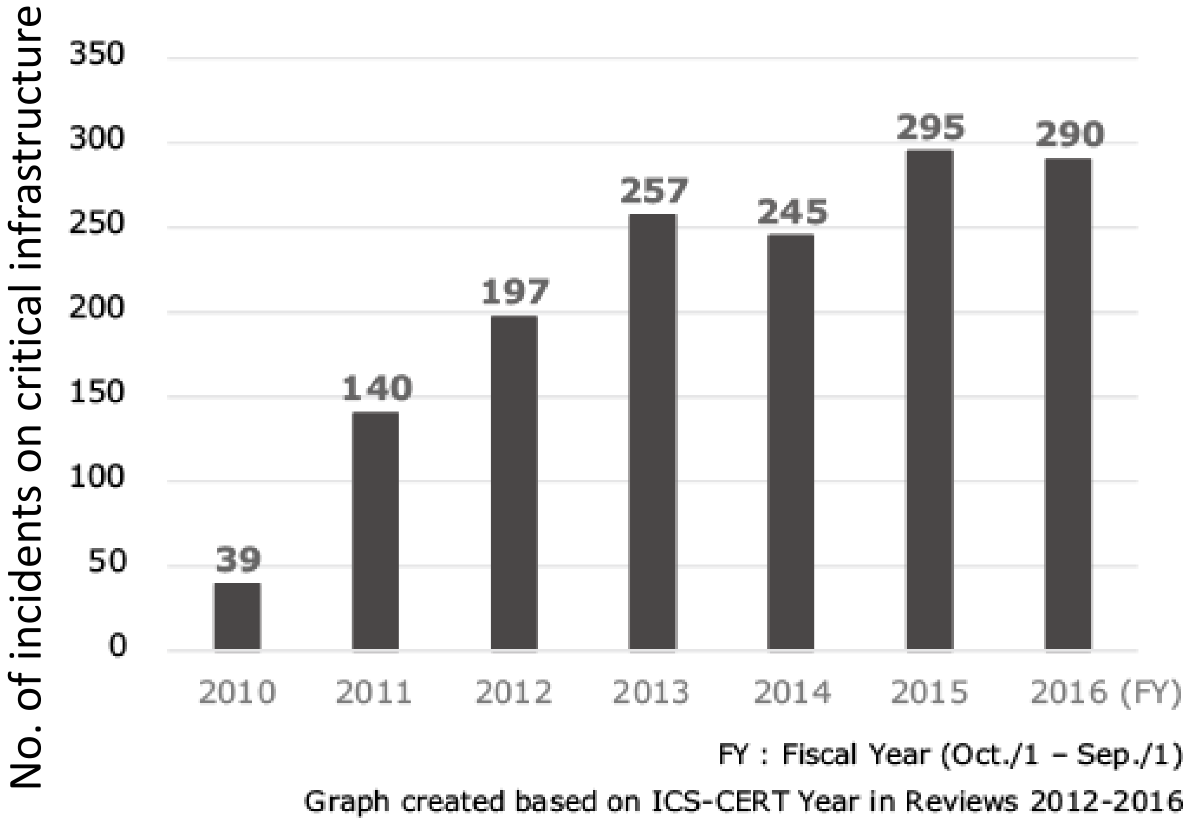


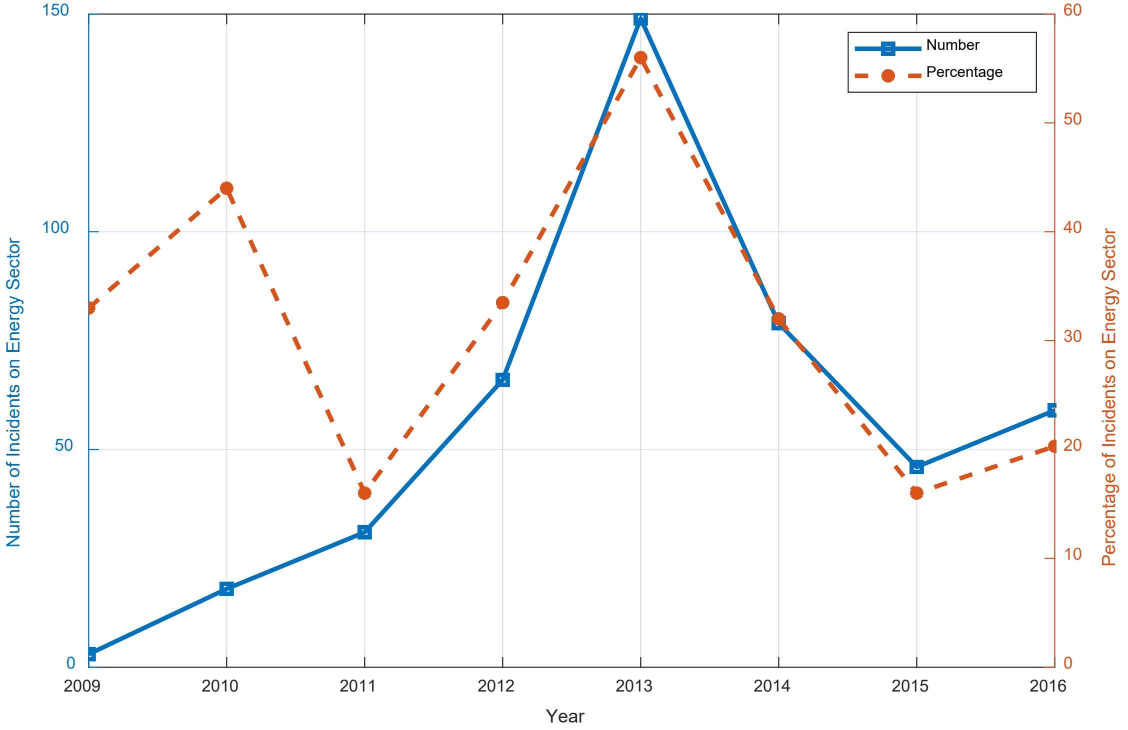


- Electric power system is vulnerable to cyber attacks via multiple POC's
  - Stuxnet malware attacks SCADA systems in Germany in 2010
  - Dragon fly attack on North American Energy Companies in 2013
  - Ukraine power grid attacks in 2015
  - Gundremmigen (German nuclear power plant) in 2016
  - Cisco Router Exploitation Kit in 2020

Need for cyber security



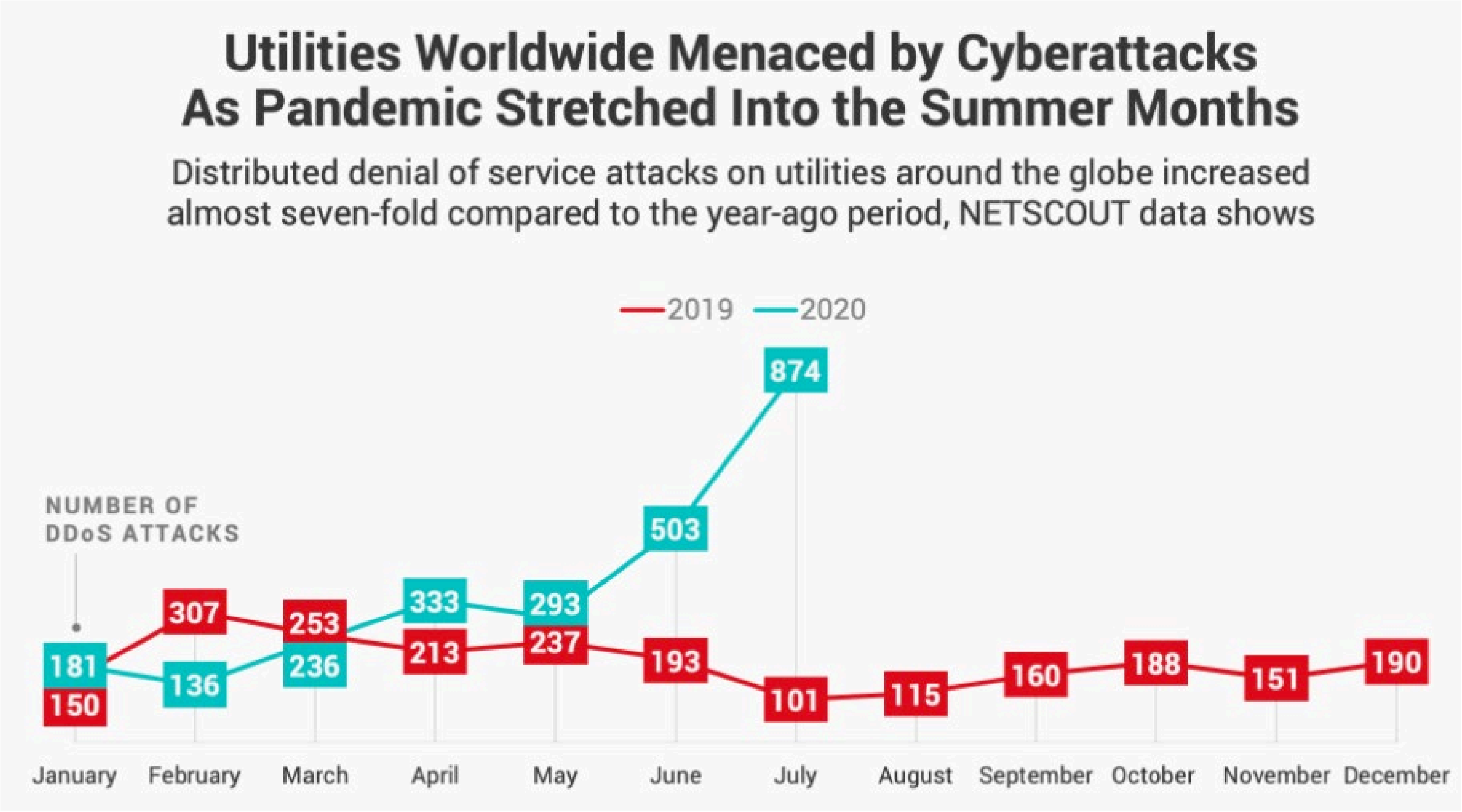




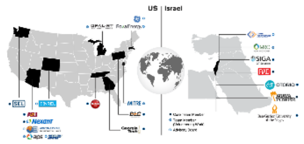
DHS recorded cyber-incidents on the energy sector [1]

[1] DHS, "ICS-CERT Year in Review Reports," [Online] Available: https://www.us-cert.gov/ics/Other-Reports

Impact of COVID-19



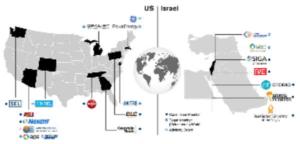
Tasks: Motivation and Overview



Two types of attacks:
- Information technology (IT) systems attack and breaches
- Operational technology (OT) systems attacks

Two types of attacks:
- Severe consequences for grid operations if monitoring and eventually control is compromised
- If IT security is ever breached (as has happened), crucial to protect grid operations from cascading failures
- Identify feasible and effective cyberattacks and defense mechanisms – contribute to attack signature knowledge base of Task 4

Two tasks:
- Task 5: Generate event-mimicking attacks
- Task 8: Detect event-mimicking attacks



Task 5: GANs for Generating Adversarial Attacks

Objective: Study the robustness and <u>fundamental limitations of detectors</u> in detecting events such as line trips, voltage dips, faults, etc

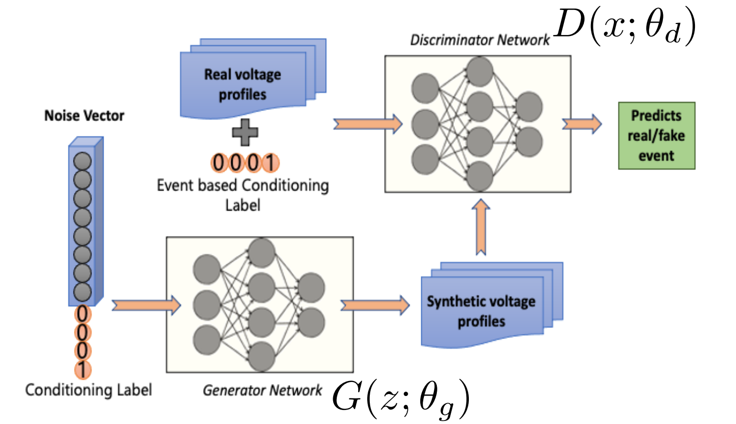
Limits of current technology:

- Completely rely on expert knowledge (supervised)
- Mostly ad-hoc without any <u>theoretical</u> guarantees or <u>rigorous testing</u> on practical datasets

Our Method: Generate adversarial examples using artificial neural network (un-supervised) based method called <u>Conditional GAN</u>. We can *generate:*

- Oscillations
- Equipment failure
- Targeted events/attacks

CGAN generating adversarial voltage examples

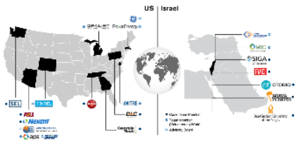


Optimization framework for CGAN

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_r(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]$$

Player 1: Discriminator **Player 2: Generator**

Task 5: GANs for Generating Adversarial Attacks



Key Highlights:

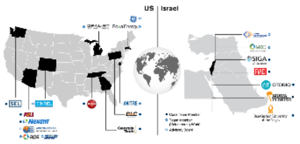
1) *Interpretability*: Although data-driven, our methods will be completely interpretable, and will aid the situational awareness of the operator
2) *Temporal Dynamics*: Develop conditional GAN architectures that considers *temporal and spatial* dependencies of the PMU data

Research Impact

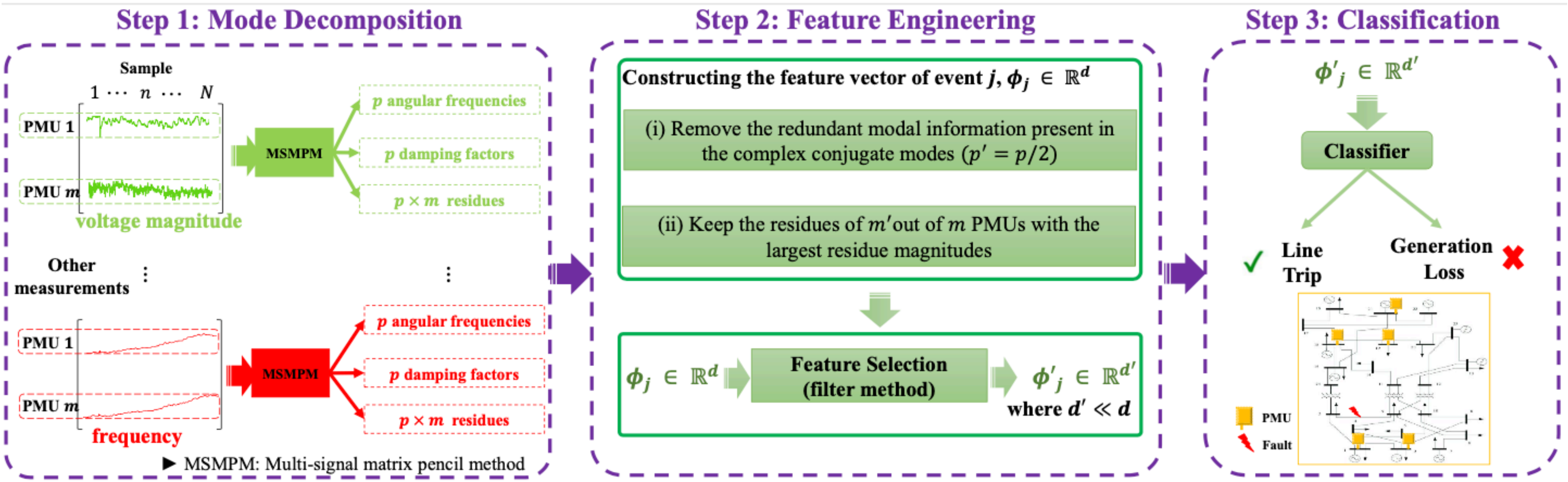
1) Harness data science to study the *fundamental limitations of existing detectors*—events or attacks—using large scale PMU data*
2) Develop *new modular software* to study the performance of detectors
3) To *incorporate the learned* "realistic" attacks to the existing knowledge base

* Publicly available data and proprietary data (if possible)

Task 5 (a): Identifying/Learning Event Signatures



- First step to generating physically realizable attacks: design accurate event detectors

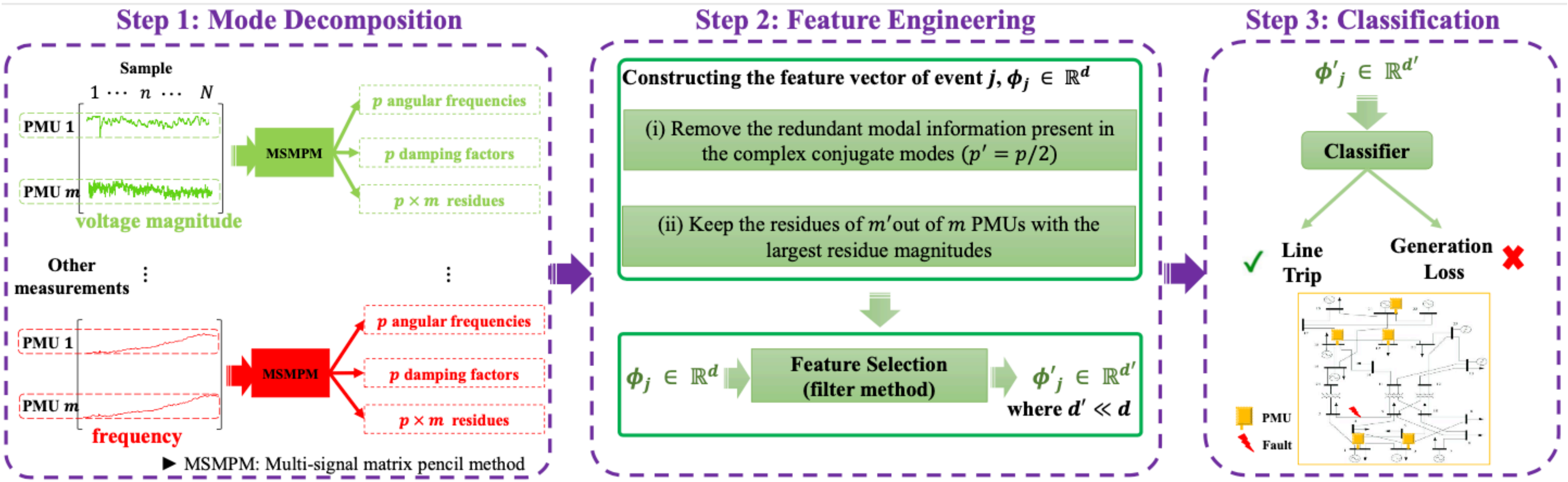


- First quarter product: Real time event identification [3]:
  - Learned features that capture physics (modes) using limited labeled data
  - Our results outperformed conventional signal processing methods widely used in industries

[3] N. Tahipourbazargani et.al (2022) A Machine learning framework for event identification via modal analysis of PMU data, submitted IEEE PES.

Task 5 (a): Identifying/Learning Event Signatures

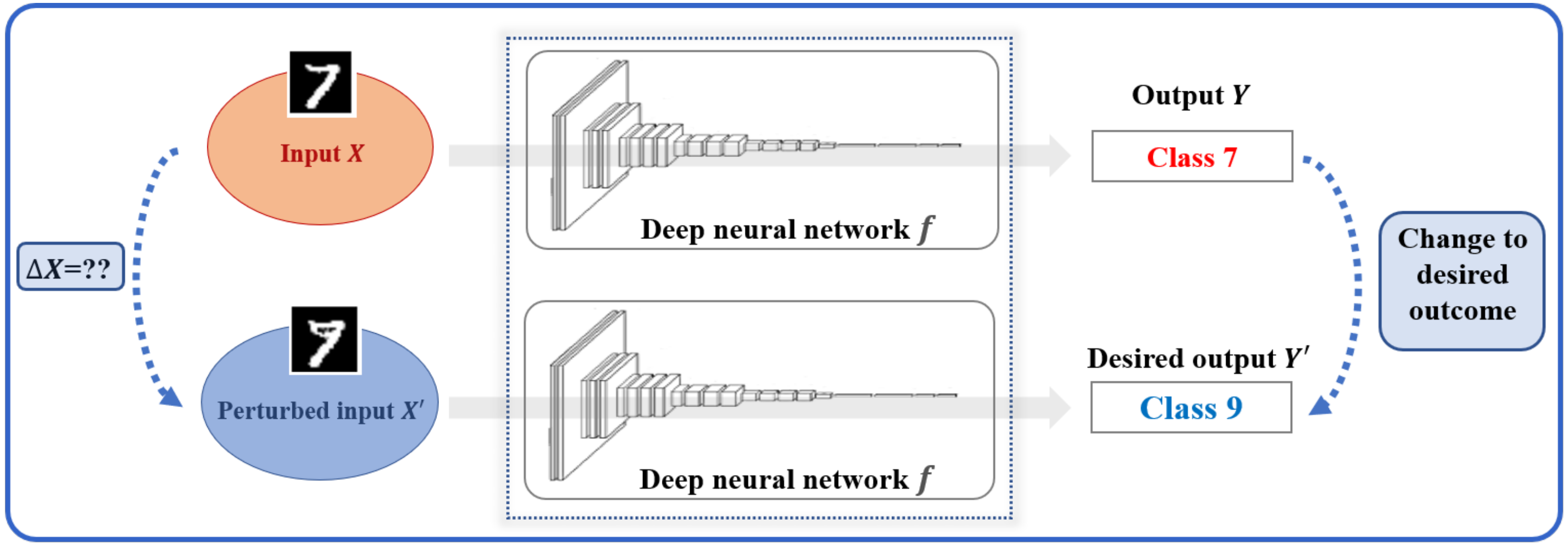
- Can we manufacture physically realizable attacks (e.g., event-mimicking)?



Yes! By identifying features that are easy to synthesize by changing measurements

Task 5 (b): Interpretable Models for Attack Generation

- Counterfactual machine learning models:

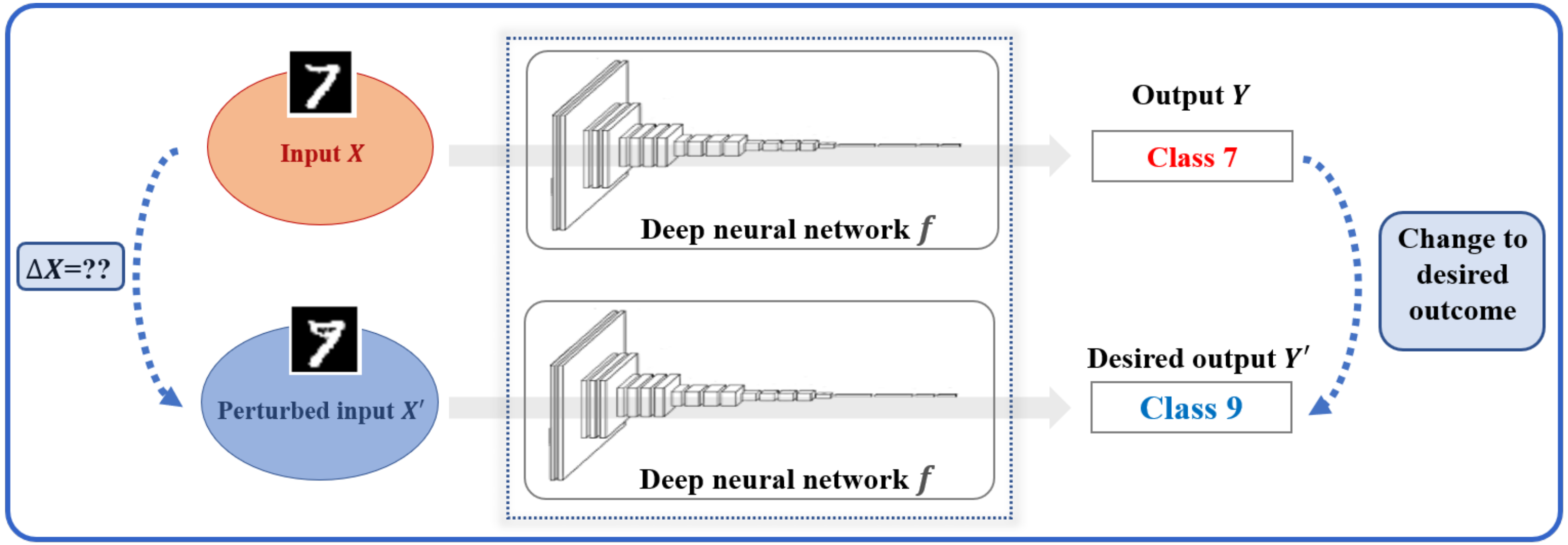


Framework of counterfactual explanation*

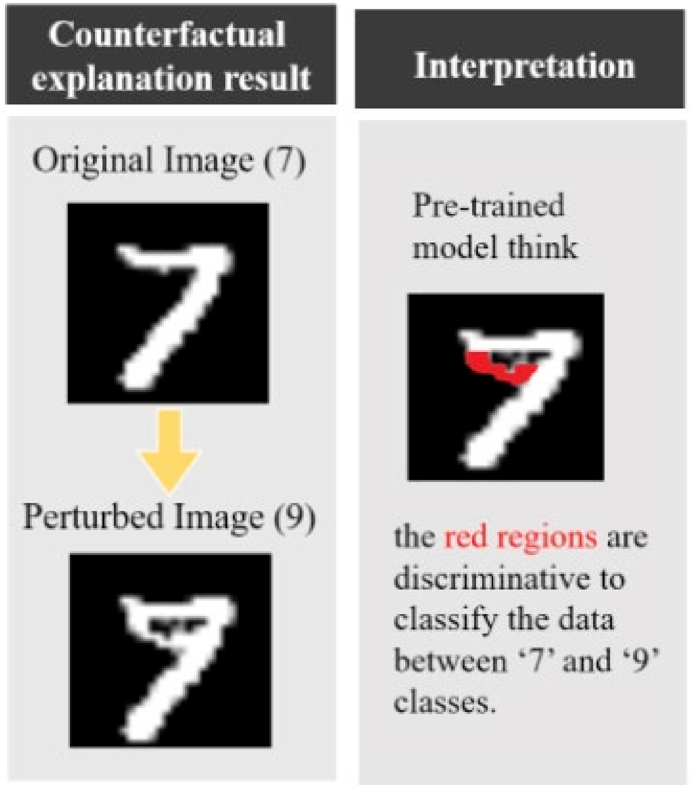
*[Online] Available: https://da2so.github.io/2020-09-14-Counterfactual_Explanation_Based_on_Gradual_Construction_for_Deep_Networks/

Task 5 (b): Interpretable Models for Attack Generation

- Counterfactual machine learning models:



Framework of counterfactual explanation*

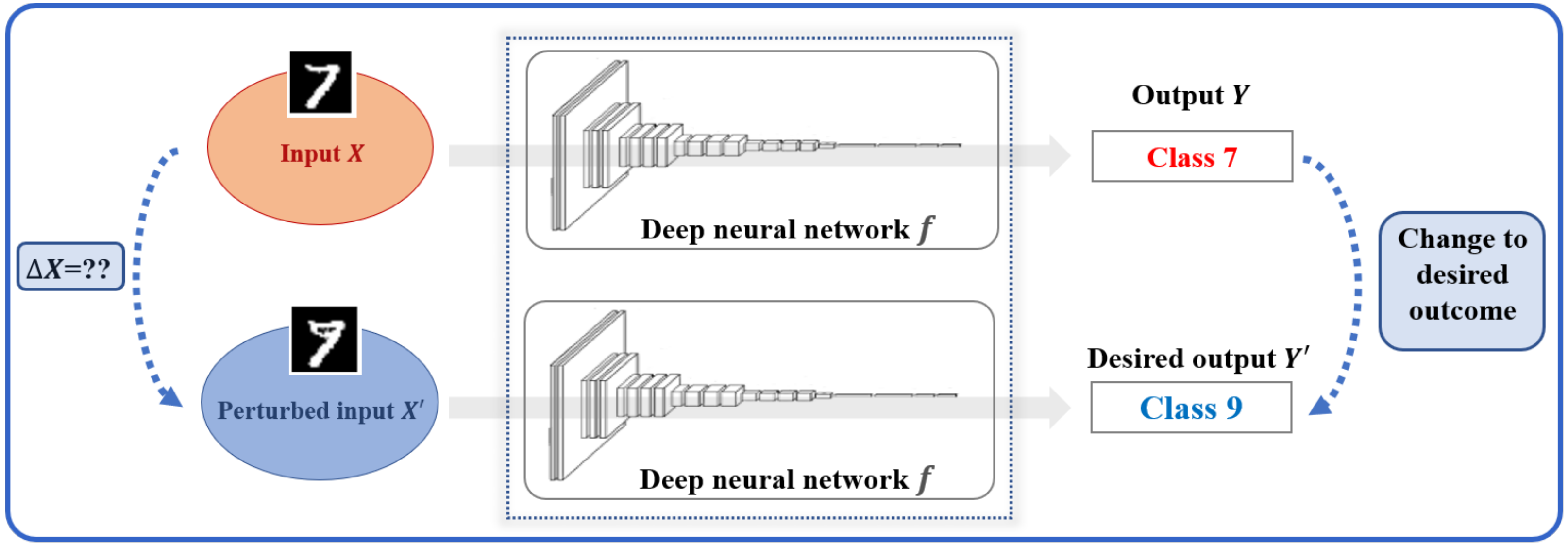


Interpretation: pre-trained detector (classifier) thinks the perturbed regions as the discriminative features between the output and desired output

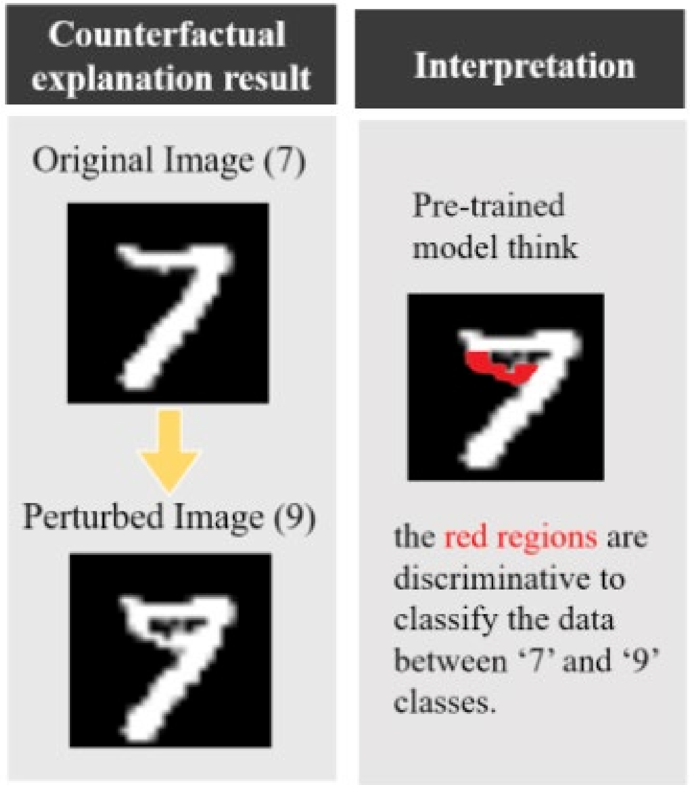
*[Online] Available: https://da2so.github.io/2020-09-14-Counterfactual_Explanation_Based_on_Gradual_Construction_for_Deep_Networks/

Task 5 (b): Interpretable Models for Attack Generation

- Counterfactual machine learning models:



Framework of counterfactual explanation*



Counterfactual models for attacks on power system attacks:

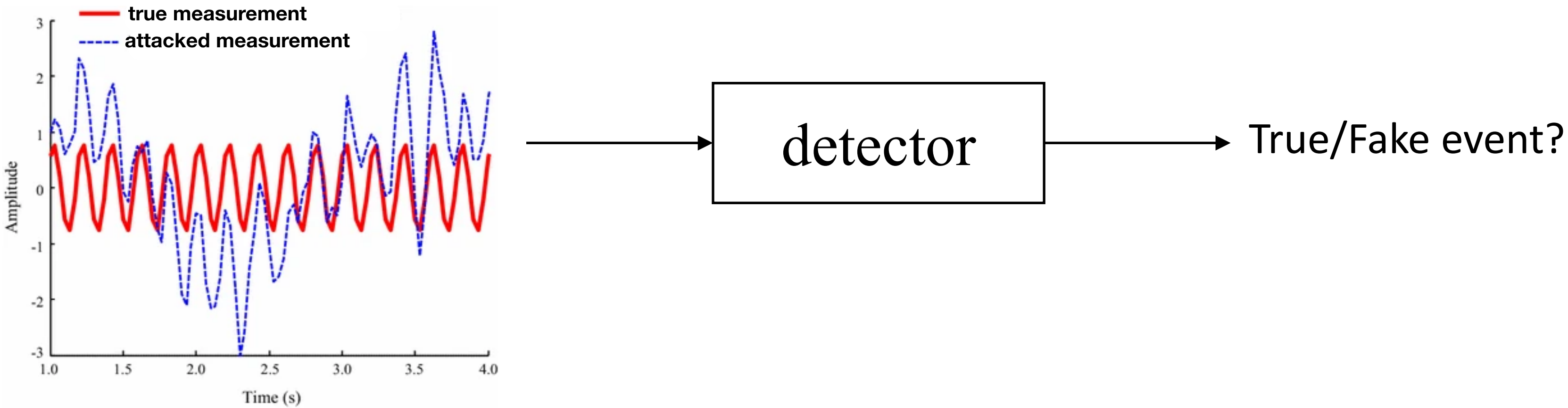
- Determine *minimal set of features with large attack impact*
- Features should be *realizable by perturbing measurements*

*[Online] Available: https://da2so.github.io/2020-09-14-Counterfactual_Explanation_Based_on_Gradual_Construction_for_Deep_Networks/

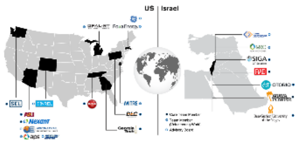
Task 8: Detect Event Mimicking Attacks

- A lot of work and software exist for developing robust detectors of static data

- How to tackle correlated time series data of (dynamical) power system ?



Task 8: Detect Event Mimicking Attacks



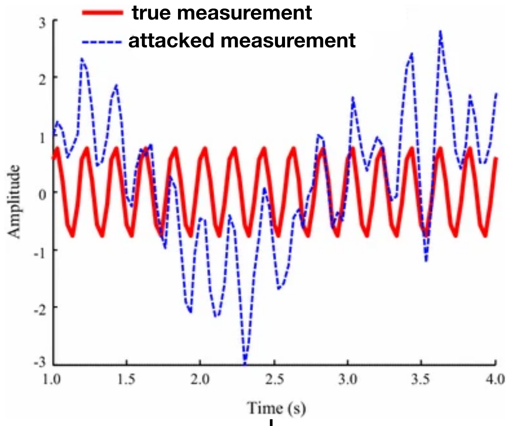
Objective: Enhance the performance of EMS by designing <u>modular detectors capable of detecting anomalies</u> via measurements

Limits of current technology:

- Model based detectors <u>do not</u> consider dynamics in PMU data
- Relies on offline/batch processing of SCADA measurements

Our Method: Online ML detector that exploits **event features** to:

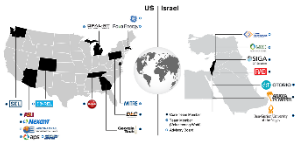
- Compare *feature signatures* of true events against fake events
- Incorporate (*physics-based*) *knowledge* to make detectors robust
- Include *event characteristics* (e.g., frequency, source of event) to enhance distinguishability



detector

True/Fake event?

Task 8: Detect Event Mimicking Attacks



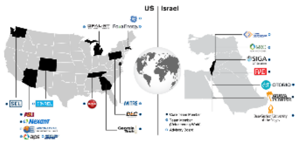
Key Highlights:

1) *Attack Space*: Mimicking attacks subsumes the existing replay and load redistribution attacks and enlarges the attack hypothesis space—*important to develop countermeasures*
2) *Modular software*: Detectors can be easily integrated/dis-integrated into the existing EMS platform without interrupting the grid operation

Research Impact:

1) Going beyond simple replay attacks to study more *realistic yet practical event mimicking attacks* by leveraging the strength of machine learning methodologies.
2) Develop detection schemes that *can intelligently fuse SCADA and PMU* measurements, thus significantly improving the detection performance
3) Quantify the performance improvement using rigorous *theoretical analysis and experimental evaluations*

Team Members and How They Interact



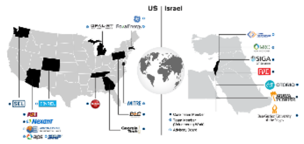
Students and postdocs:

- *Nima TagizhpourBazarghani (grad student)*: Design interpretable and physics-based machine learning methods to detect and classify events
- *Obai Bahwal (grad student)*: Design, implement, and evaluate GAN (adversarial) attacks and event mimicking attack detectors on real and synthetic data sets
- *Dr. Rajasekhar Anguluri (postdoc):* Develop rigorous theoretical guarantees for the above frameworks and mentor the grad students

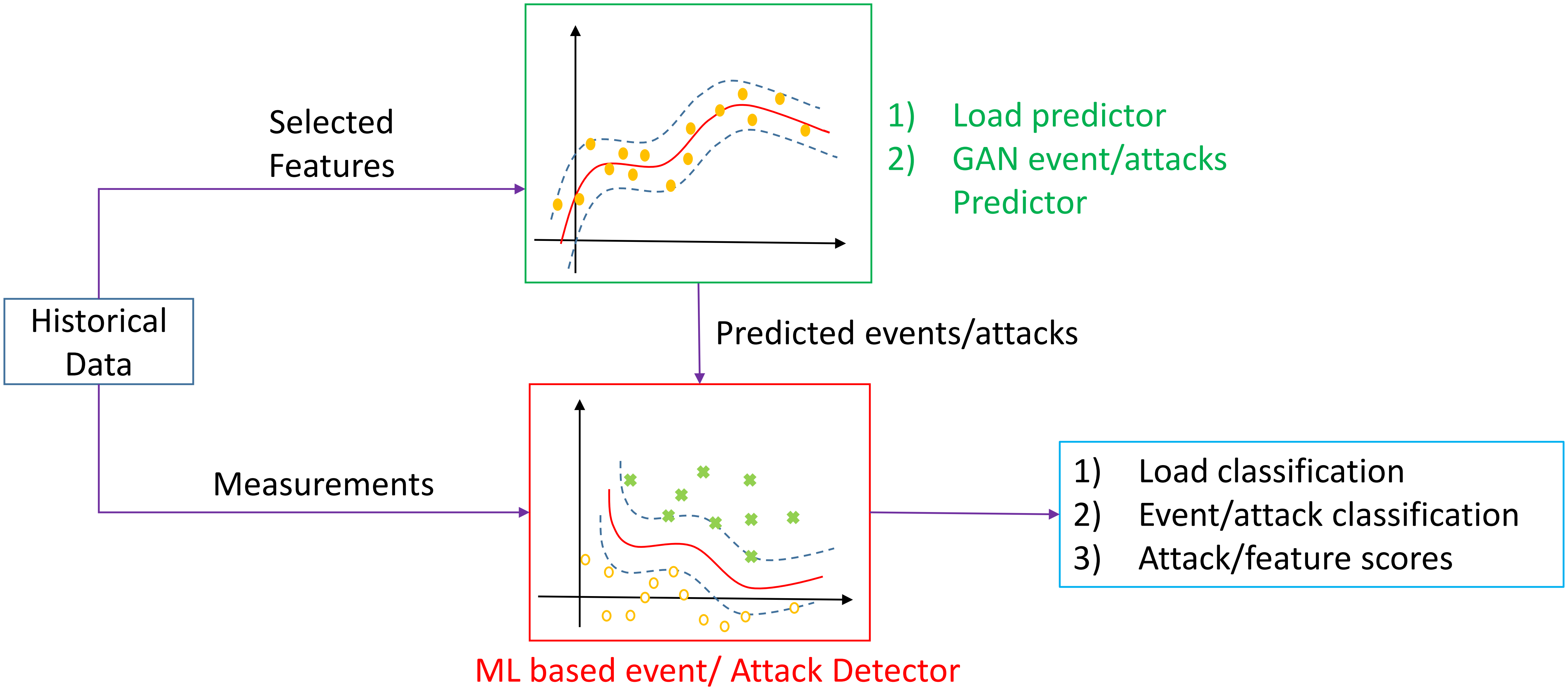
Research Collaborators:

- *Dr. Yang Weng (Asst. Prof., ASU, and BIRD project lead)*: Discuss how the modular *outcomes of Tasks 5 and 8 can enable next-generation EMS*
- *John Dirkman, Nexant*: Collaborate to *secure and enhance Nexant's Grid360 tool*
- *Dr. Oliver Kosut (Assoc. Prof, ASU)*: (On-going collaborations of Sankar) Collaborate on event classification work via Kosut's expertise in optimization and cyber security

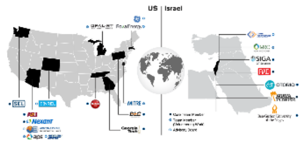
Commercialization Plan

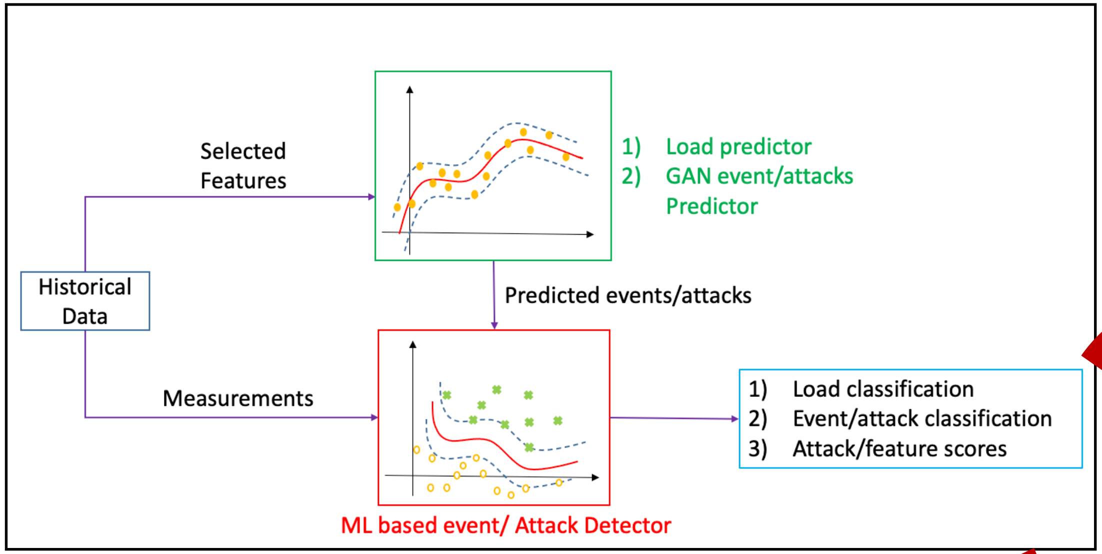


- **New module**: takes in monitoring data and evaluates its authenticity including attacks, events, faults, to name a few



Commercialization – From Detection to Anomaly Visualization





Intelligent and interpretable attack/event detector

Nexant Grid360: Load Anomaly Visualization