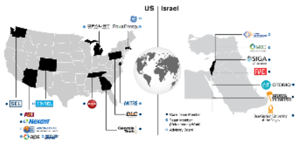
Event-mimicking Attacks and Countermeasures



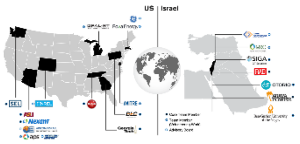
Task 5: Generate event-mimicking attacks

Task 5(a): mimic modal features of PMU data

Task 5(b): From lab to practice (Nexant/RRI)

Task 8: Detect event-mimicking attacks

ASU Team Members



Obai Bahwal
1st Year PhD Student

Rajasekhar Anguluri
Postdoctoral Fellow

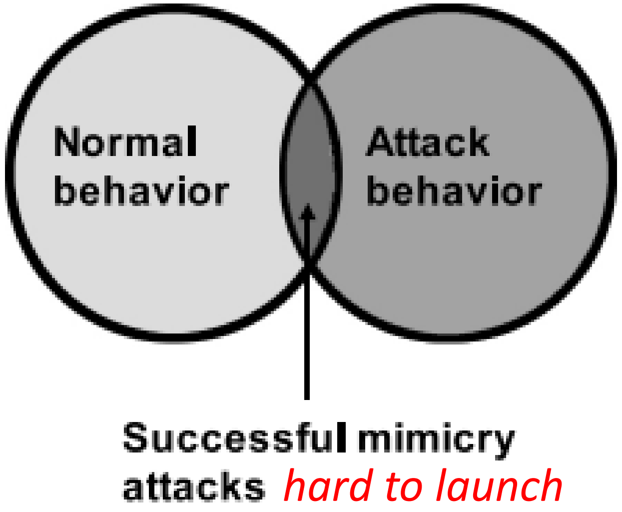
Joel Mathias
Postdoctoral Fellow

Nima T. Bazargani
4th Year PhD Student

Also collaborating with John Dirkman, Fernando Magnago, Suresh Babu Argi @ Resource Innovations Inc.

Event-mimicking Attacks and Countermeasures

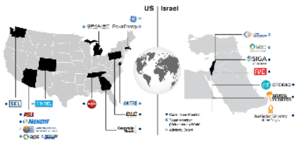
- Modern grid with renewables is more stochastic in operations and requires real-time monitoring to detect/identify real events (oscillations/outages) and attacks.
- ML-based detectors can be easily evaded by attacks that mimic events, ultimately, causing significant damage on grid operations.

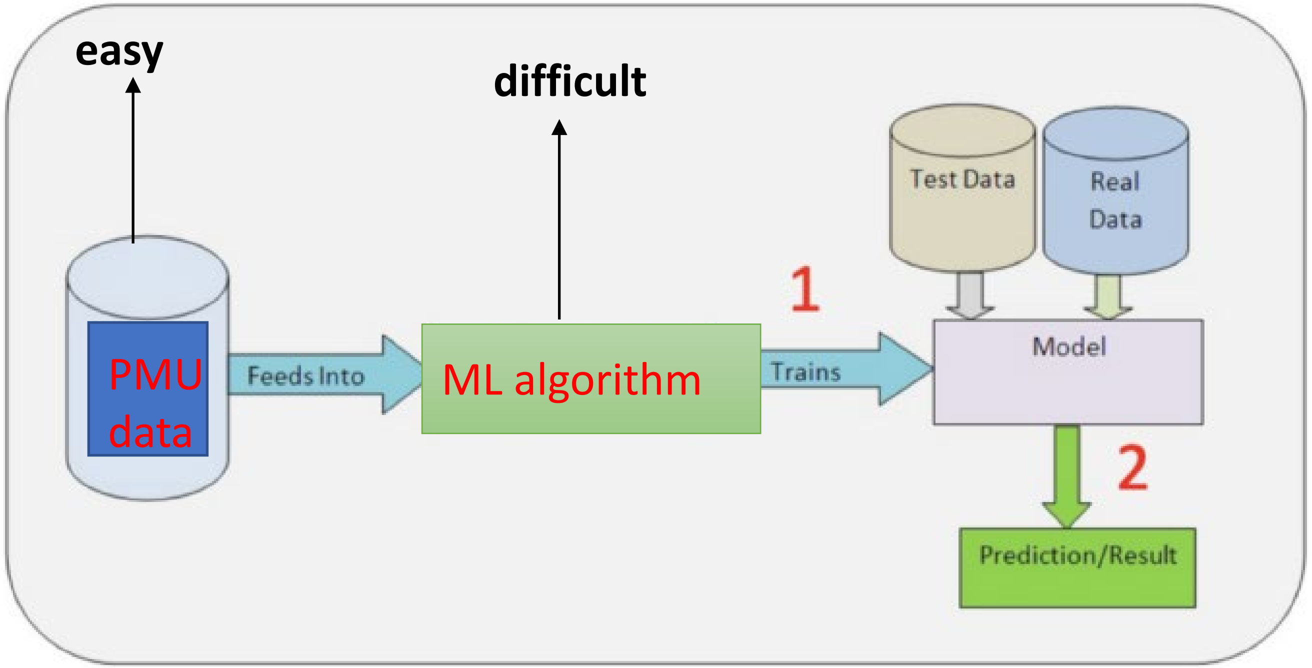


mimicry attack: a careful cyberattack on data that throws off ML detector

Source: https://towardsdatascience.com/evasion-attacks-on-machine-learning-or-adversarial-examples-12f2283e06a1

Where can Attackers target in OT Systems?





easy to tamper PMU ; but for mimicking event attacks

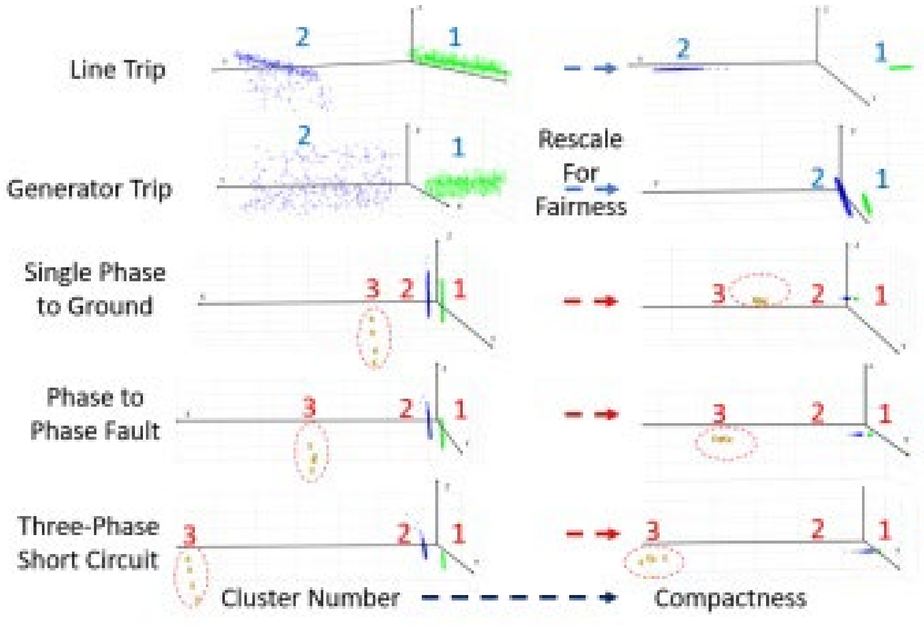
- *how to tamper data?*
- *how many PMUS to tamper?*
- *how long to tamper?*

extract and exploit signal physics (modes)

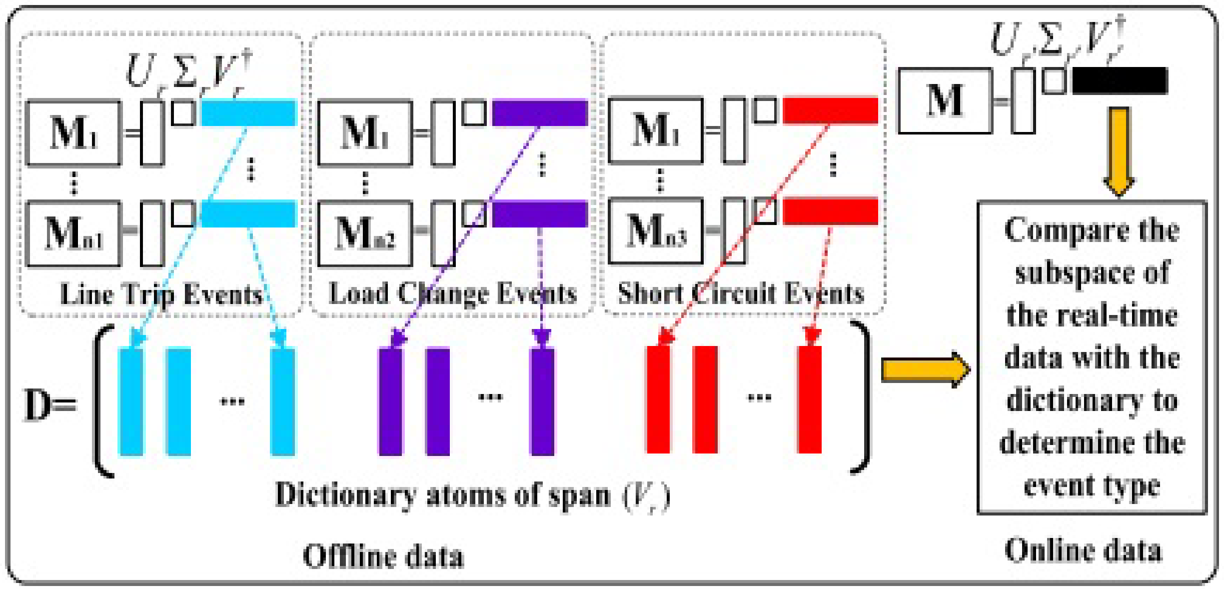
Source: https://towardsdatascience.com/will-my-machine-learning-be-attacked-6295707625d8

Task 5 (a): Learn Event Signatures from Measurements

Prior work neglects the physics (e.g., modes, residues, frequency) encoded in PMU data



Unsupervised learning for event detection

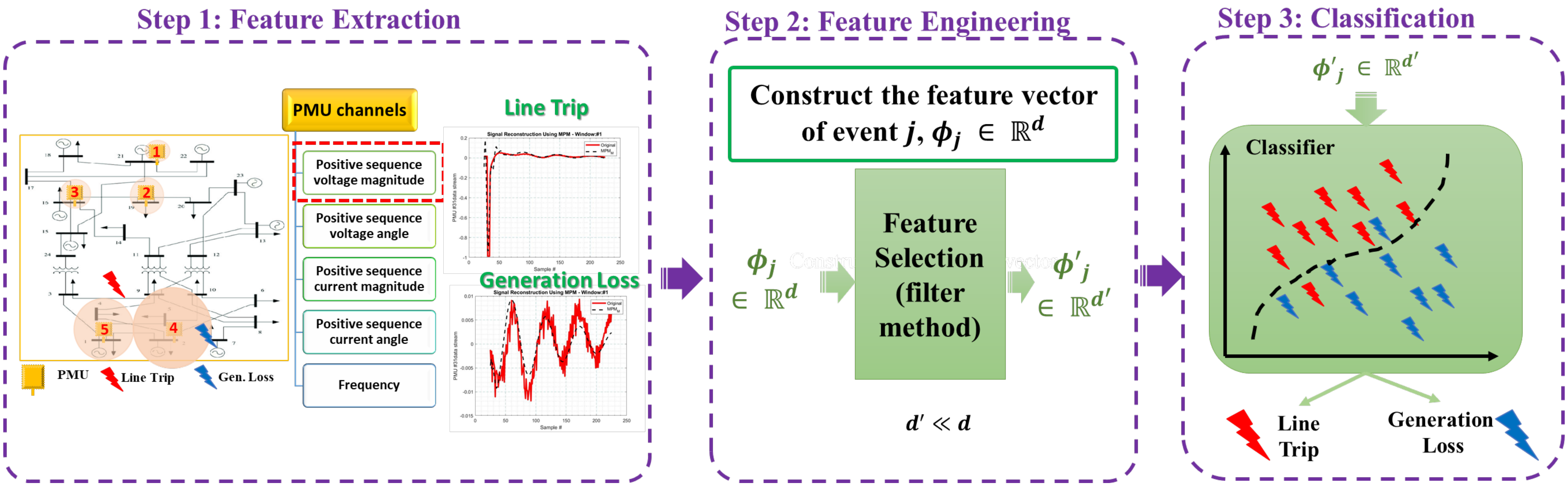


Dictionary based learning for event identification

H. Li, et. al, "**An Unsupervised Learning Framework for Event Detection, Type Identification and Localization Using PMUs Without Any Historical Labels**," *PES GM 2019*

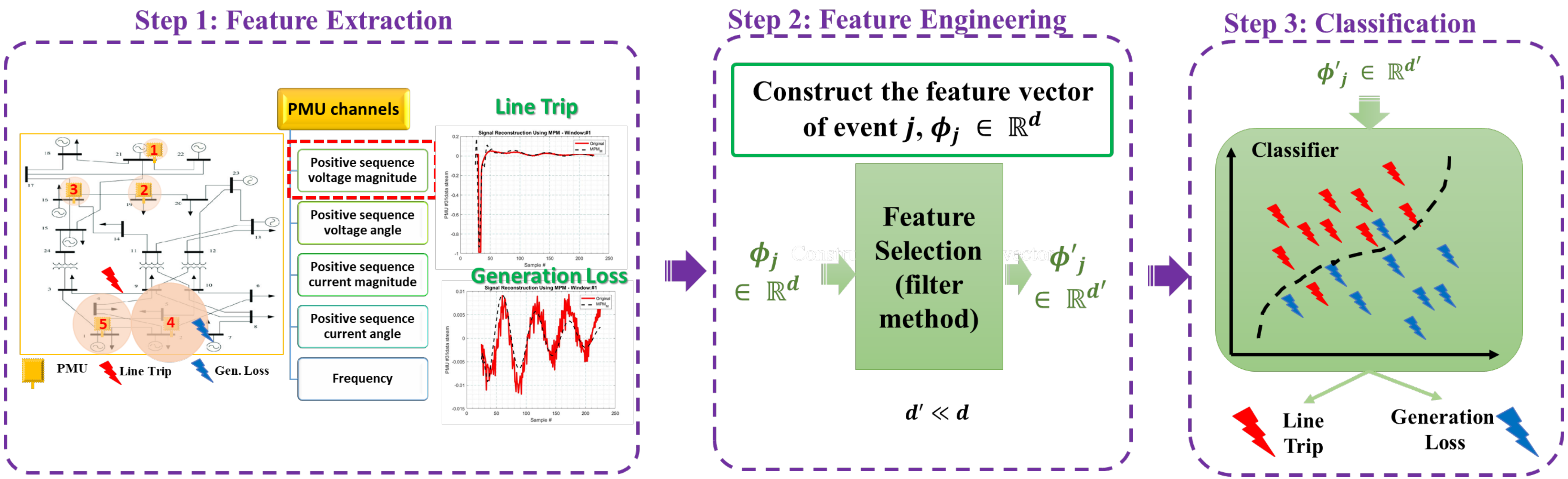
W. Li, et. al., "**Real-Time Event Identification Through Low-Dimensional Subspace Characterization of High-Dimensional Synchrophasor Data**," *IEEE TPS*, vol. 33, no. 5, pp. 4937-4947, Sept. 2018.

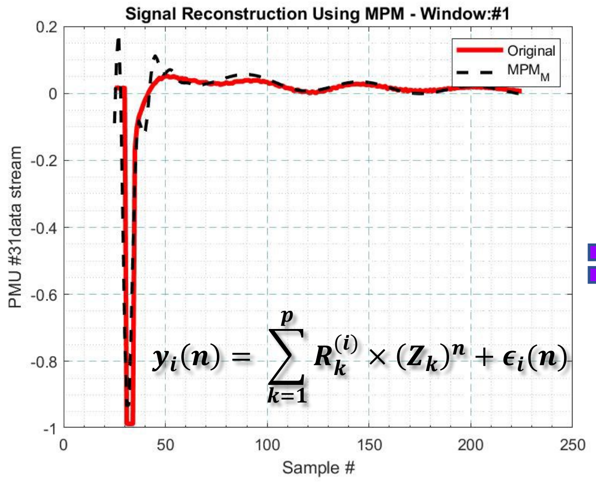
Task 5 (a): Learn Event Signatures from Measurements

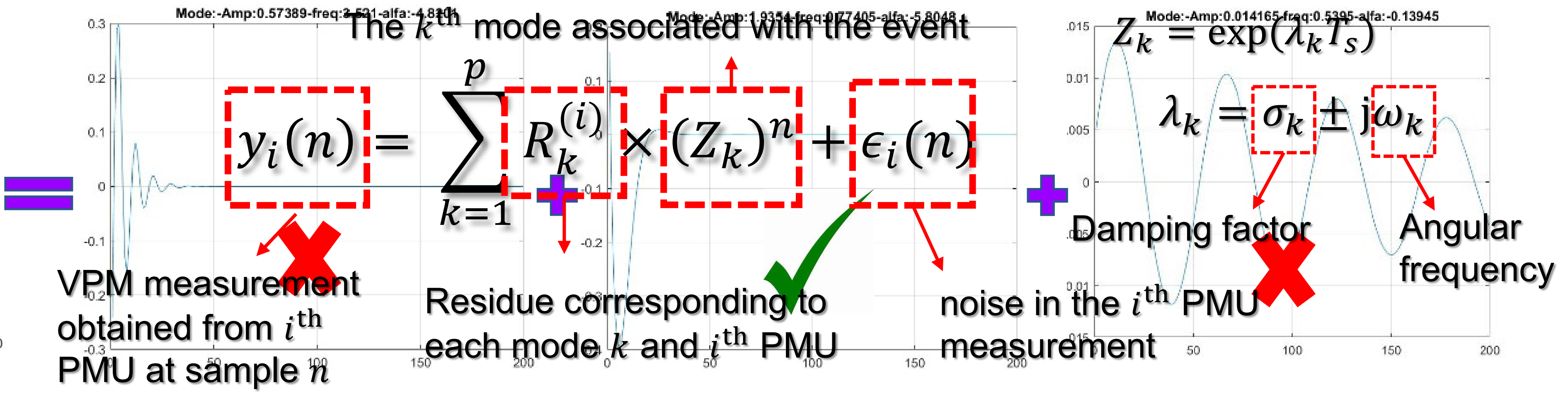


- ✓ Characterizing events based on a set of physically interpretable features
- ✓ Finding the most informative sparse set of features
- ✓ Learning a set of robust classification models to identify the events

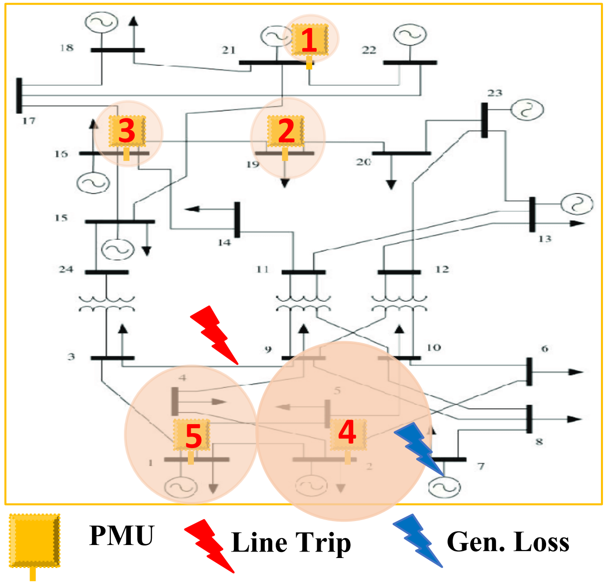
Task 5 (a): Learn Event Signatures from Measurements

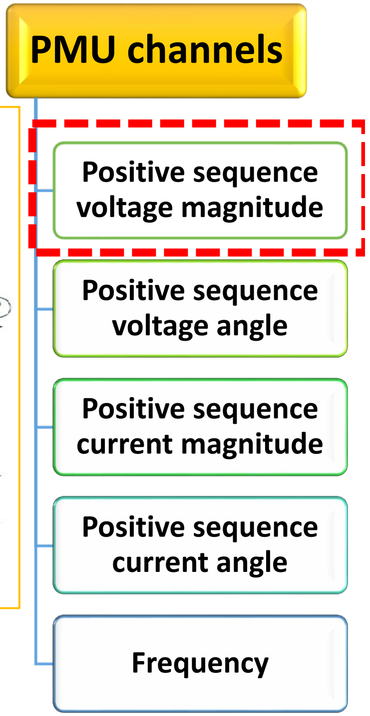


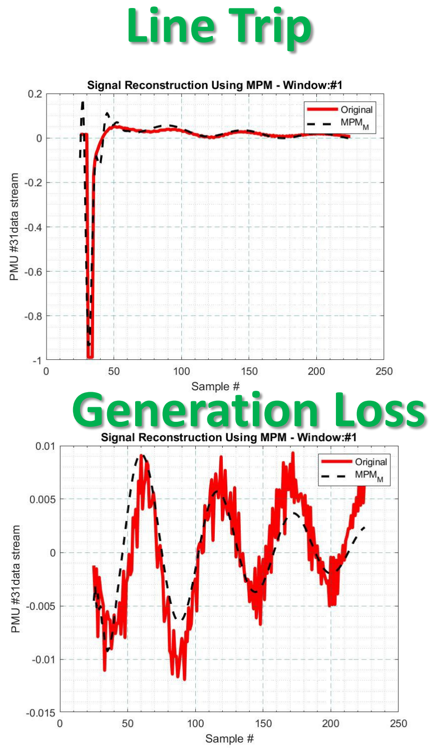




Task 5 (a): Learn Event Signatures from Measurements







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

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

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

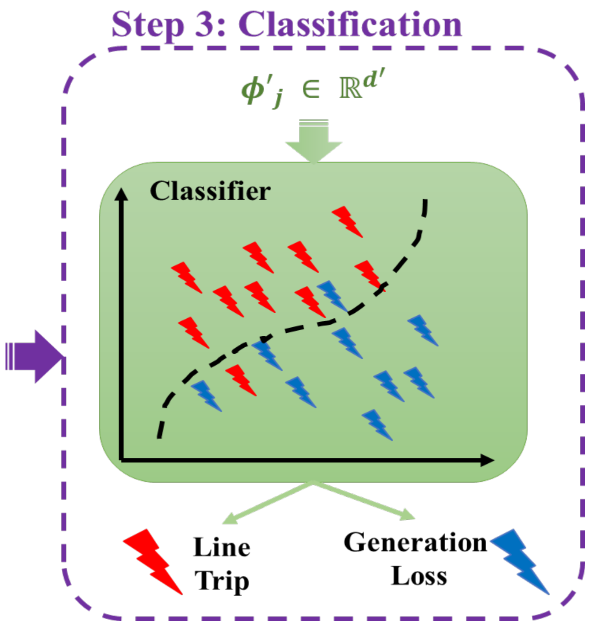
Task 5 (a): Threat Model

- (start with) White Box Attack Model: Attacker has full information of the event classifier
- Untampered Features: $\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$

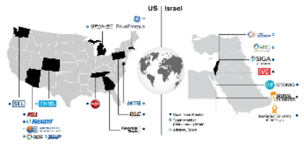
$$\boldsymbol{y_i(n)} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{p}} \boldsymbol{R_k^{(i)}} \, (\boldsymbol{Z_k})^{\boldsymbol{n}} + \boldsymbol{\epsilon_i(n)}$$

$$\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$$

\# of features: $p' < p$
\# of PMUs: $i = 1, \dots, m' < m$
$ch = Vm, Va, F$



Task 5 (a): Threat Model

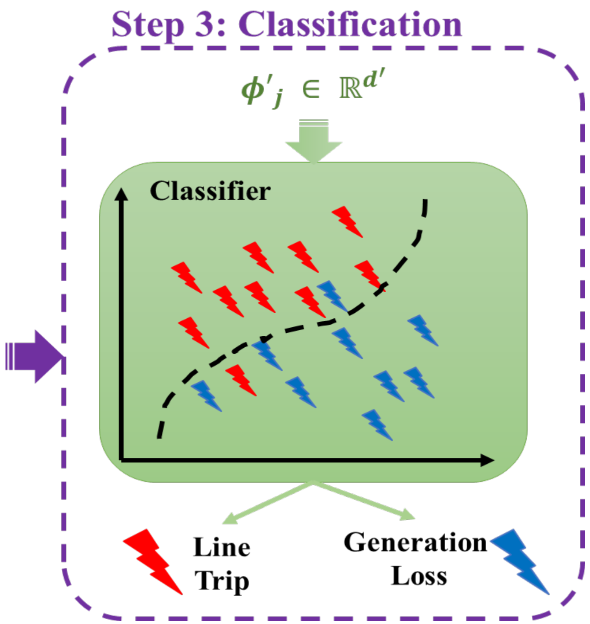


- **Start with White Box Attack Model:** Attacker has full information of the event classifier
- Untampered Features: $\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$
- Which features can be tampered for maximal impact / misclassification?
  - not your usual additive false data injection

$$\boldsymbol{y_i^{atk}(n)} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{p}} \boldsymbol{R_k^{(i),atk}} \times \left(\boldsymbol{Z_k^{atk}}\right)^{\boldsymbol{n}} + \boldsymbol{\epsilon_i(n)}$$

$$\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$$

\# of features: $p' < p$
\# of PMUs: $i = 1, \ldots, m' < m$
$ch = Vm, Va, F$



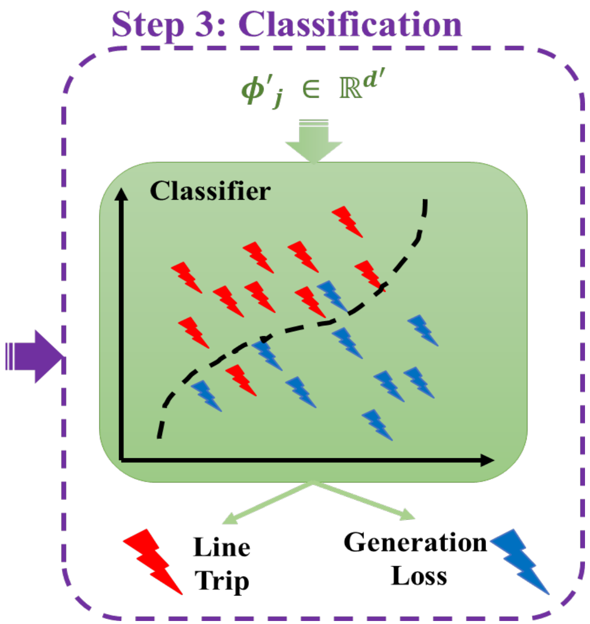
Task 5 (a): Threat Model

- Start with White Box Attack Model: Attacker has full information of the event classifier
- Untampered Features: $\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$
- Which features can be tampered for maximal impact / misclassification?
- First attack effort: **tamper with residual amplitudes** $\mathcal{F}_{\text{ch}}^{\text{ATK}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|^{\text{ATK}}\right\}, \{\theta_k^{(i)}\}\right]$

$$\boldsymbol{y_i^{atk}(n) = \sum_{k=1}^{p} (R_k^{(i)} + x) \times (Z_k)^n + \epsilon_i(n)}$$

$$\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \left\{\left|R_k^{(i)}\right|\right\}, \{\theta_k^{(i)}\}\right]$$

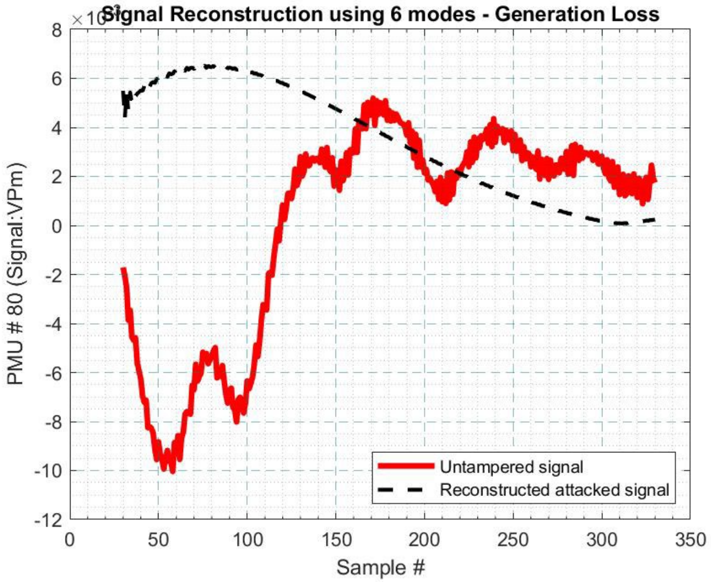
\# of features: $p' < p$
\# of PMUs: $i = 1, \ldots, m' < m$
$ch = Vm, Va, F$

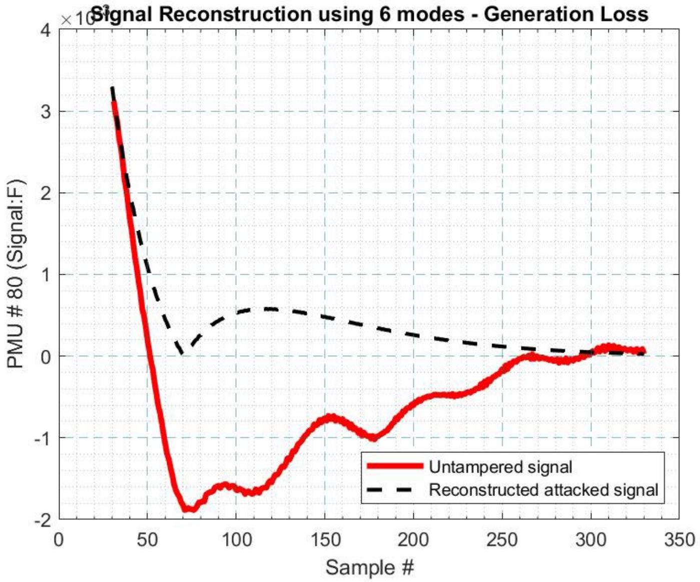


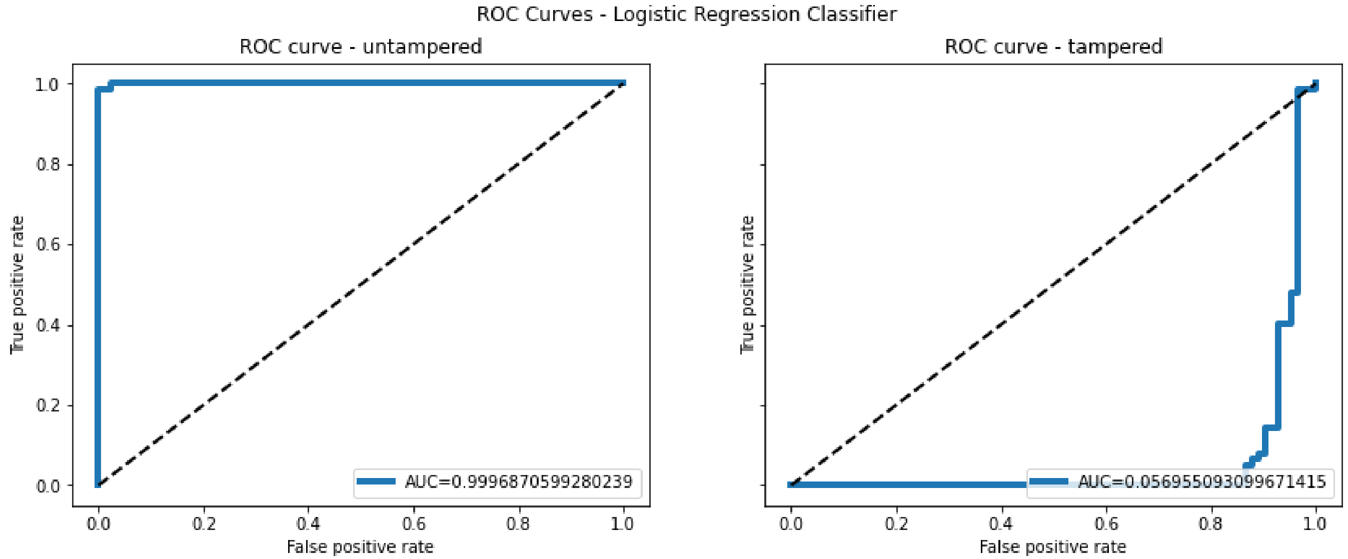
Task 5 (a): Tampering Residual Magnitudes

- Tampering technique: add <u>20 to all first mode residues</u> for both Vm and F channels
- Illustration shown here for **Generation Loss** event
- Similar results for Line Trip events can be shown

- AUC/ROC curves show that misclassification is possible
- However, time-series signal has too large an amplitude and could potentially be detected as anomalous (simple energy-based anomaly detection could work)







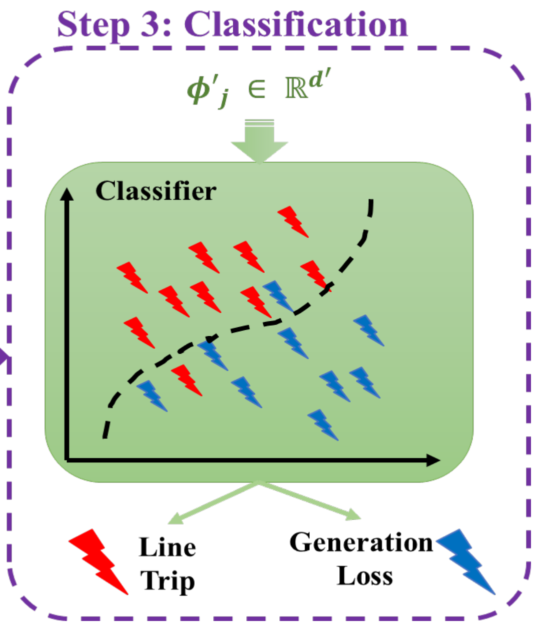
Task 5 (a): Threat Model

- Start with White Box Attack Model: Attacker has full information of the event classifier
- Untampered Features: $\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \{\left|R_k^{(i)}\right|\}, \{\theta_k^{(i)}\}\right]$
- Which features can be tampered for maximal impact / misclassification?
- 2nd attack effort: **tamper residual angles:** $\mathcal{F}_{\text{ch}}^{\text{ATK}} = \left[\{\omega_k\}, \{\sigma_k\}, \{\left|R_k^{(i)}\right|\}, \boxed{\{\theta_k^{(i)}\}}^{\text{ATK}}\right]$

$$\boldsymbol{y_i^{atk}(n) = \sum_{k=1}^{p} (R_k^{(i)} + x) \times (Z_k)^n + \epsilon_i(n)}$$

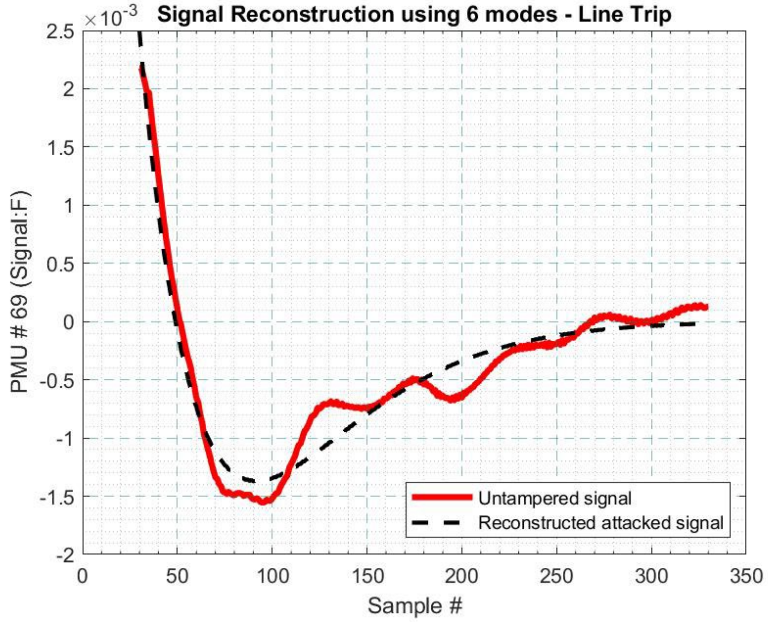
$$\boxed{\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \{\left|R_k^{(i)}\right|\}, \{\theta_k^{(i)}\}\right]}$$

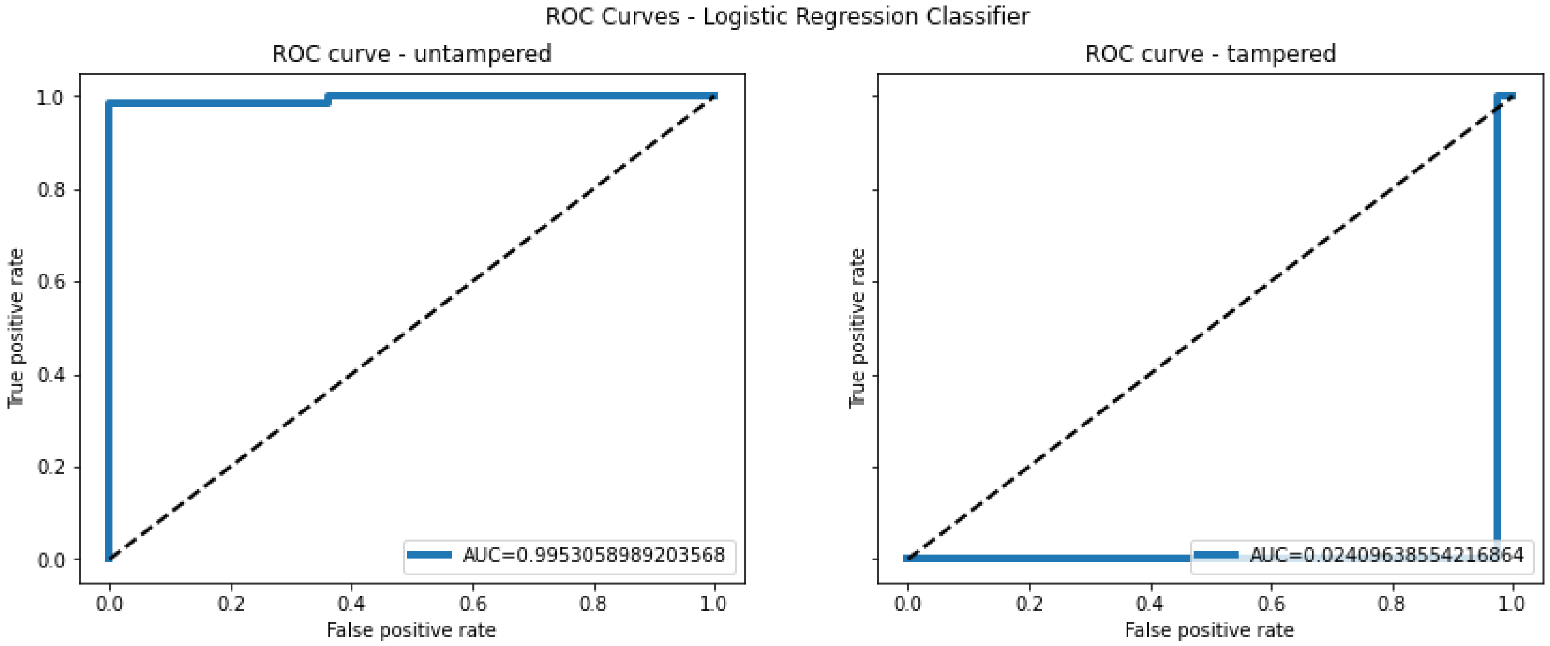
\# of features: $p' < p$
\# of PMUs: $i = 1, \ldots, m' < m$
$ch = Vm, Va, F$

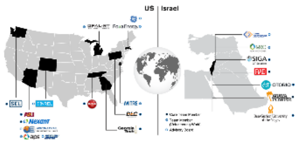


Task 5 (a): Consequences of Tampering Residual Angles

- Tamper residual angle of the first three PMUs with highest residues
- Angles modified by adding 100π
  - succeeds in spoofing the classifier
  - reconstructed signal indistinguishable from original
- However, attacker needs to tamper classification algorithm to spoof features directly







Task 5 (a): Threat Model

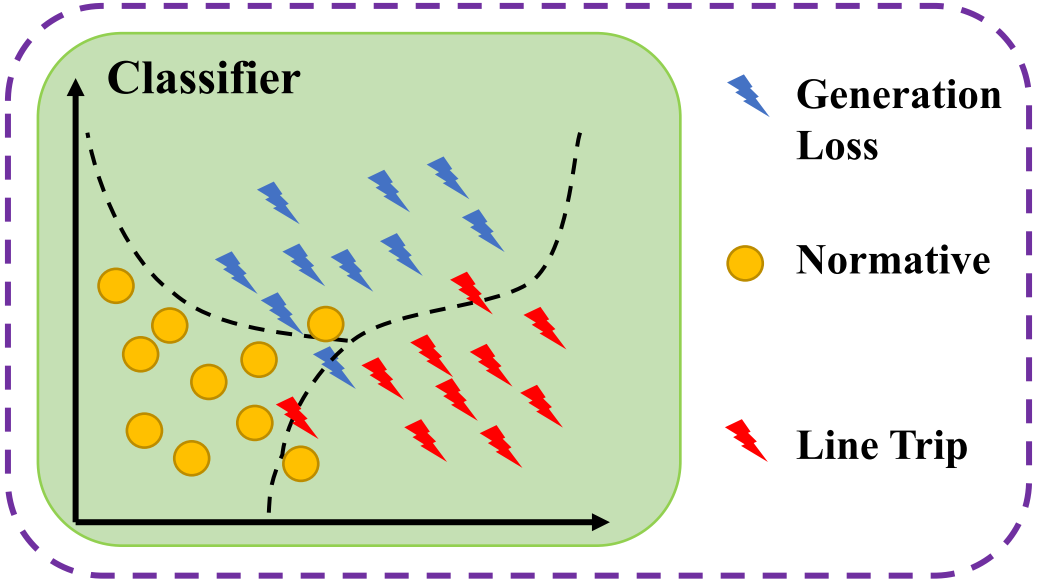
- White Box Attack Model: Attacker has full information of the event classifier
- Untampered Features: $\mathcal{F}_{\text{ch}} = \left[\{\omega_k\}, \{\sigma_k\}, \{\left|R_k^{(i)}\right|\}, \{\theta_k^{(i)}\}\right]$
- Initial tests: tamper residual amplitudes and angles – either need large values or more access

Attack avenues being explored:

- Can we intelligently tamper modes $\{\omega_k\}, \{\sigma_k\}$: key signatures of an event?
- How can topology information be utilized to identify most susceptible PMUs?
- Attacks are expensive and identifying a small set of features and PMUs to attack is crucial

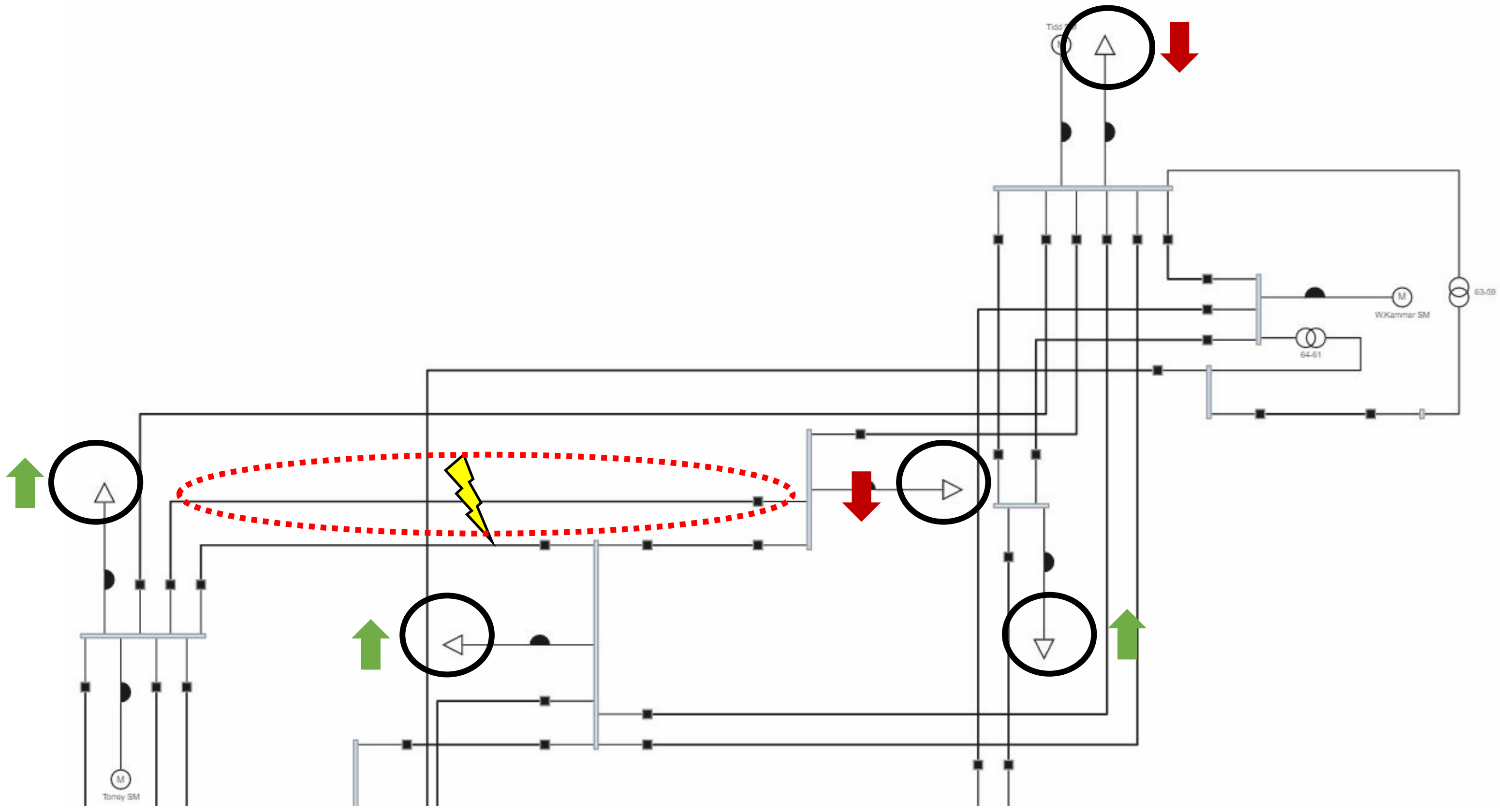
Task 5 (a): Next Steps

- System largely operate in normative conditions
- However, when events occur, they can be of more than two types
- A natural extension to multi-event classification is to include a third normative class (non-event class)
- Are attacks easier (even white box ones) in the multi-class setting? We conjecture: yes
- Key challenge: designing intelligent attacks without resorting to brute force – requires exploiting physics of the data without breaking physical laws

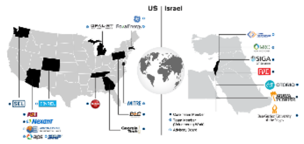


Task 5 (a): Collaboration with RII

- Industry Collaboration: Resource Innovations, INC (RII)
- Attack design on RII's Grid360 power flow simulator
- IEEE118 sub-transmission network model is used
- Loads modified on network sub-region such that net change is zero
- **Goal:** cause line overflow undetectable by conventional state estimators



Task 5 (a): From Lab to Practice



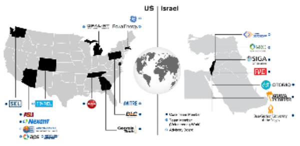
- Industry Collaboration: Resource Innovations, INC
- Exploring Grid360's capabilities to understand where our research fits in
- Understanding the intelligence of the state estimator by varying loads measurements and checking for bad data flags
  - Working on RI team on multiple bugs that were discovered
  - In the process of being fixed – weekly on-going meetings with RI

SE is able to detect a load change of 10 MW

Device Type	Measurement Type	Measurement Value	Estimate	Error	RDFID
GEN	MW	155.0000	158.1162	-3.11621	_4E270DE60274AD6E8B0F7EDA82464BAF
LOD	MW	342.4000	335.5162	6.88380	_8989B6A8D764C187160EDEEB666B7C00
GEN	MW	338.3569	338.3569	-0.00003	_BD6DAFE655FFF6E9D8441B16A7AF076C
GEN	MVAR	408.6065	408.6065	0	_C853A528180C3473FAACCCB5B2651268

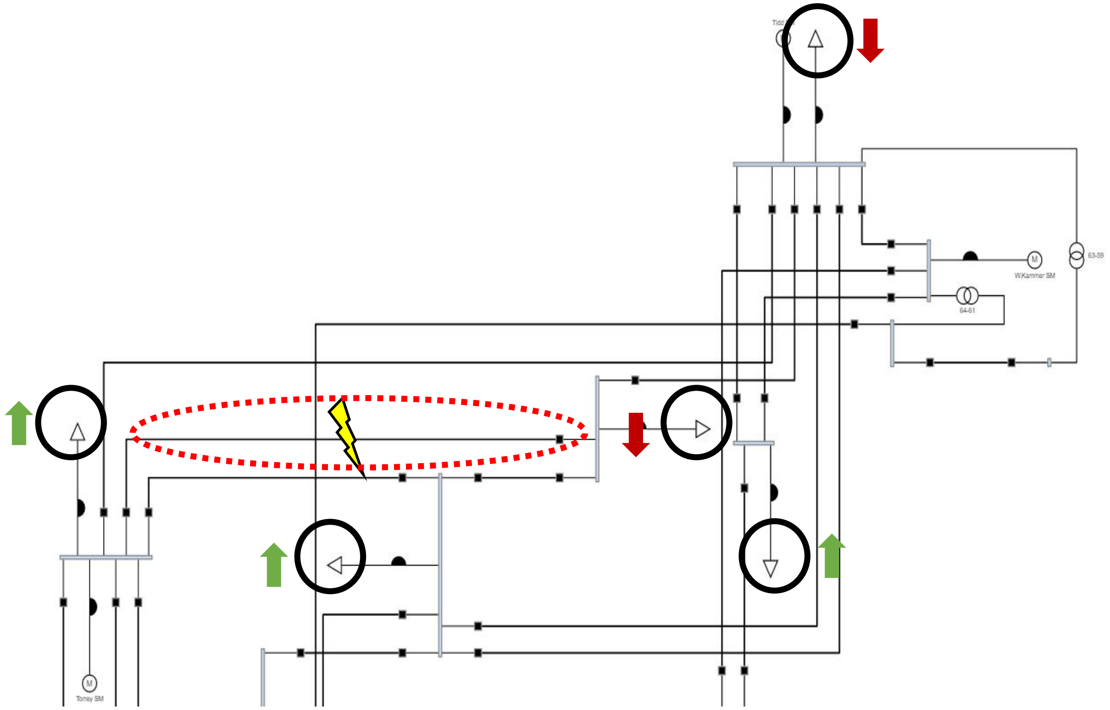
SE failed to detect a load change of 1 MW

Device Type	Measurement Type	Measurement Value	Estimate	Error	RDFID
GEN	MW	338.3569	338.5297	-0.17275	_BD6DAFE655FFF6E9D8441B16A7AF076C
GEN	MVAR	408.6065	408.4812	0.12531	_C853A528180C3473FAACCCB5B2651268



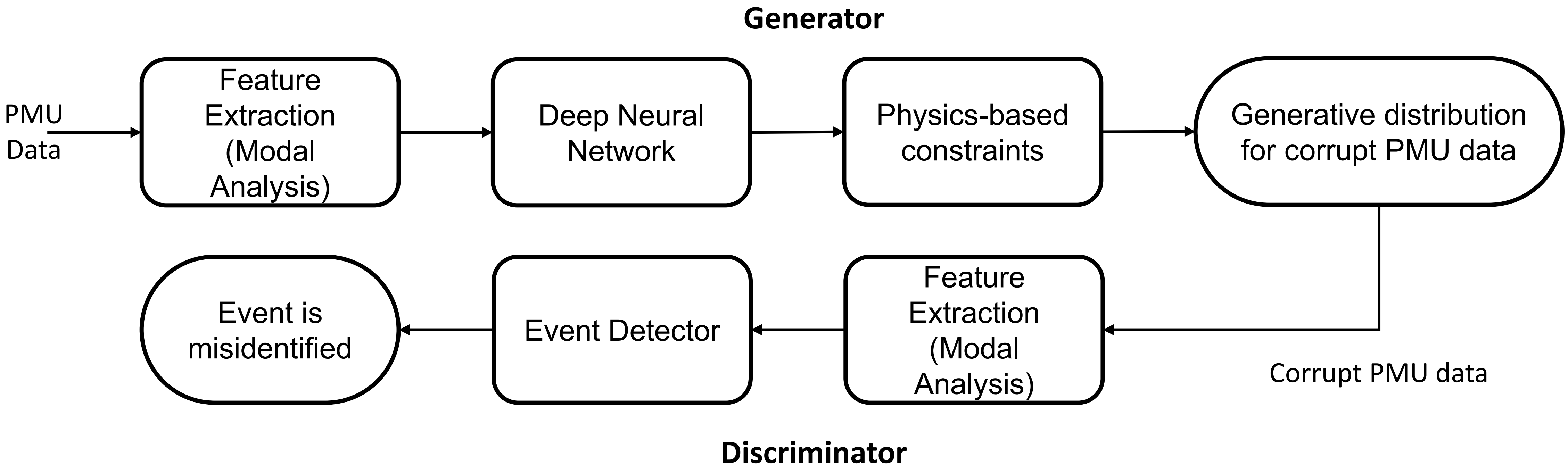
Task 5 (a): Commercialization by RII

- **Evaluate efficacy of load attacks on RII's Energy Management System platform**
- Use SCADA data and simulate on Grid 360 software
- Can their conventional state estimator detect an attack?
- Counter measures: use sophisticated machine learning techniques to improve state estimation under attacks
  - Flag anomalous loads that results from false measurements injected
- Work closely with RII as they test our robust EMS algorithms (e.g., bad data detector) towards commercialization
  - Key idea: use tomes of history data+ML

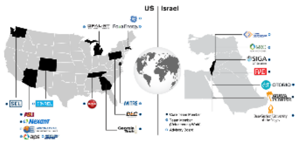


Task 5 (a): Future directions

- A Deep Learning framework for attacks
- Learn generative model of corrupt PMU data
- Utilize knowledge of feature extraction process and physics of signal
- Adversarial training of generator: detector spoofed into misidentifying events



Summary of Work for Q3



	Details	Status
Task 5 (attack generation)	• **Synthesize “intelligent” attacks that mimic “events” by Tampering measurements.**	• **Completed feature extraction** • **Analyzing features realizable by altering measurements.**
Task 8 (attack detection)	• develop ML and data-driven “robust” detectors that detect intelligent attacks.	• In two quarters.
Industry Collaboration	• **Seamlessly integrate ML detector to Nexant Grid360 tool.**	• **Pilot study: test our prior load-altering attacks and detectors using ”smart-meter” data.** • **Towards product: in four quarters.**