CPSIoTSec 2020 Virtual Event @ Orlando USA

Towards Robust Power Grid Attack Protection using LightGBM with Concept Drift Detection and Retraining

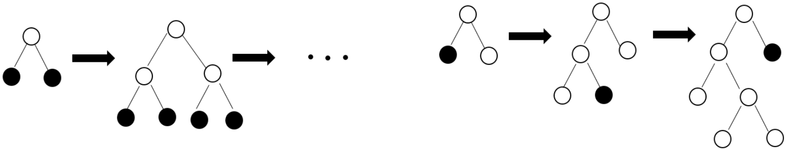
ANAND AGRAWAL MARIOS SAZOS AHMED AL DURRA MICHAIL MANIATAKOS

Overview

- In this work we present a robust methodology to detect cyber attacks on the Power Grid
- Issues with the state-of-the-art methods:
  - Model-based approaches lack scalability and accuracy
  - Focus mostly on binary (attack/no attack) events, ignoring natural events
  - Machine learning-based model accuracy reduces overtime
- In this work:
  - We employ a realistic testbed to expose all attack surfaces of the platform
  - We use a Three-Class classifier (no attack/natural event/attack)
  - We introduce concept drift to maintain the model accuracy in the long-term

Proposed Base Classifier: LightGBM

- Ensemble Learning
  - Multiple learners are trained in sequential or parallel way
  - LightGBM: Light Gradient Boosting Machine
    - LightGBM excludes data instances with small gradients
    - Converges faster and accurately



Threat Model

- Assumptions
  - Attacker has control over monitoring data
  - Compromised devices may be protection/monitoring or networking
- Examined attack scenarios:
  - Data injection: Change phasor values in synchrophasor packets
  - Remote tripping injection: Sending close/trip command using relays
  - Relay setting change: Change the configuration of relay i.e. CT, VT ratio

Training Dataset

- Benchmark Dataset[1]
  - Consists of 3 bus system and 4 circuit breaker
  - Buses equipped with relays to control the circuit breakers
- Dataset description
  - *Scenarios*: Dataset consists of 37 scenarios
  - *Events*: Dataset broadly categorized as No Events, Natural, and Attack Events
  - *Features*: Comprises of 128 features correspond to phasor and magnitude of voltage and current

Feature	Description	Importance
R2-PA6:IH	Relay2 Current Phase Angle	0.020763
R2-PA3:VH	Relay2 Voltage Phase Angle	0.019436
R3-PA6:IH	Relay3 Current Phase Angle	0.019231
R2-PA7:VH	Relay2 Voltage Phase Angle	0.019213
R1-PM5:I	Relay1 Current Phase Magnitude	0.019079
snort_log	Log collected by IDS	0.000000
control_panel_log	Log collected by control panel	0.000000
R2: S	Status Flag for relays	0.000000
R3-PA9:VH	Relay3 Voltage Phase Angle	0.000086
R4-PA9:VH	Relay4 Voltage Phase Angle	0.000089

Class	Description	Event Type	Scenario Id.
0	Normal operation load changes	No Event	41
1	SLG Faults, Line Maintenance	Natural Event	1-6,13,14
2	Data Injection, Remote tripping, Relay setting change	Attack Event	7-12,15-20, 21-30, 35-40

1 https://www.sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

Feature Selection

- Accuracy of reduced feature set and overall set
  - Accuracy of model trained from selected features is better
  - In multi-class accuracy increases with more features
- Other features
  - Number of iterations: 8000
  - Max_depth: 9, num_leaves: 50
  - boosting_type: Gradient Boosting (GDBT), Gradient-based One Side Sampling (GOSS)

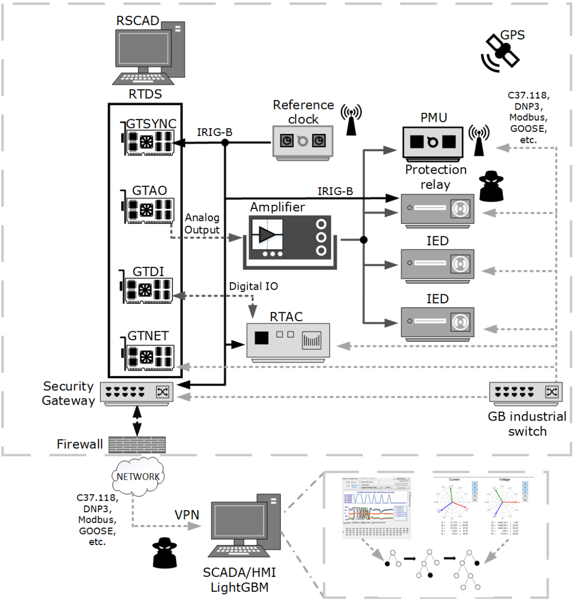
Features \ Label	Binary-Class	Three-Class	Multi-Class
All	96.9	95.3	92.65
Subset	97.3	97.1	91.2

Model Generation and Concept Drift

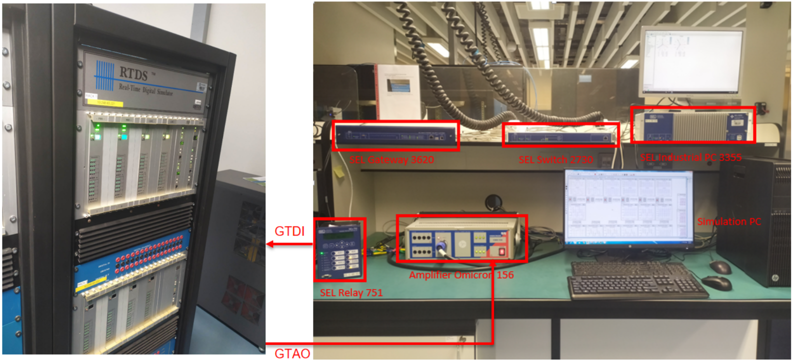
- Detect accuracy reduction
  - Divide dataset in chunks
  - Pass data to Massive online Analysis (MOA)[2] framework
  - Drift Detection Method and Early Drift Detection Method
  - Detection of drift at instances where accuracy is reduced substantially
- Retraining of model
  - LightGBM classifier is retrained based on drift detection
  - Model accuracy increases in fraction

[2] A Bifet, G Holmes, R Kirkby, and B Pfahringer. 2010. MOA: Massive Online Analysis. J. Mach. Learn. Res. 11 (2010)

Experimental Setup Schematic View



Experimental Setup Physical View



Model Detection Accuracy

- Evaluation Metrics and detection
  - Compare results against other machine learning approaches
  - Used Binary, Three, and Multi class label data

Boosting Classifier	DNN			SVM			LightGBM (Our work)		
[L] / [M]	B	T	M	B	T	M	B	T	M
Precision	68.53	61.26	10.36	83.17	79.05	20.64	97.38	**96.99**	93.11
Recall	78.16	78.26	13.48	78.57	78.57	24.34	97.38	**96.98**	92.76
F1 Score	68.86	68.72	6.01	69.44	69.43	19.05	97.38	**96.98**	92.76
Accuracy	78.16	78.26	13.48	78.57	78.57	24.34	97.28	**96.98**	92.76

[M]: Metrics, [L]: Labels, B : Binary Class, T : Three Class, M : Multi Class

- Dynamic retraining of model after drift detection
  - Retrain model after 50% accuracy loss
  - Binary classification and Three class accuracy increased to 98.02% and 97.73% respectively

Related Work Comparisons

جامعة نيويورك ابوظبي
NYU ABU DHABI

- We compare against emulated or physical power system
  - LightGBM outperforms other boosting techniques
  - Existing solutions do not employ concept drift
  - Our method performs best in three class data

Related Work	Classifier	Label	# of features	Acc.	CD
Our Work	LightGBM	T	128	97.73%	✓
[26]	Adaboost	M	128	93.54%	✗
[10]	XGboost	T	128	95%	✗
[27]	SAE DL	B	128	94.91%	✗
[22]	NNGE+STEM	M	128	93%	✗
[21]	CNN	T	150	94%	✗

SAEDL: Stacked-Autoencoder Deep Learning, NNGE+STEM: Non Nested Generalized Exemplars State Extraction Method, CNN : Convolution Neural Network, B : Binary Class, T : Three Class, M : Multi Class, CD : Concept Drift

Conclusions and Future Direction

This work-in-progress provides three main insights:

- Ability of dynamic retraining of ML based model
- Focus on three class recognition, to include natural events
- Generalize the model for any power grid setup

Future work: We are in process of collecting more data with increased feature set, and we will also consider timestamp and drift based on timing.

Q&A

جامعة نيويورك أبوظبي
NYU | ABU DHABI

Thank you for your time!