Towards more observable grid

• Event triggered measurements
  – Relays
  – Digital fault recorders
  – Power quality meters

• Continuous measurements
  – SCADA
  – AMI (advanced metering infrastructure)
  – PMU (phasor measurement unit)
  – Point-on-wave (POW) measurements

• Event records
  – Outage and maintenance records
  – Device activation records

But still one step away

• Data labeling is critical to AI/ML
  – MNIST
  – ImageNet
  – BTO Building Benchmark Datasets

• Challenges exist for grid events
  – Data is decentralized and inaccessible
    • Limits actionable data available for analytics
  – Data is multimodal and unstandardized
    • Prevents integration of different data sources
  – Data is unprocessed and unvalidated
    • Lacks critical metadata and proper labeling
Project Overview

• ORNL and LLNL, funded by DOE Office of Electricity, partnered to develop an open-source Grid Signature Library (GSL)
  – Measurement data: raw data with signatures yet to be extracted
  – Signature data: labeled events with data provided in specific formats

• Goal
  – Facilitate, tag and fuse data feeds from multiple sources
  – Implement a modular architecture for expandable design
  – Anonymize event sources to enable open data sharing
  – Provide go-to resources for event detection and algorithm validation
Library Framework

Multi-format Data Ingestion

IEEE Standards
CSV Files
Streaming data
Other

Data Standardization
Secure Storage and Organization of Structured Data

Multi-format Data Export

Time Series Database

Event Labeling

Anomalies ingested to a new database

GSL

Detected Events

AI/ML-based Grid Health Monitoring

Data modeling and validation

Data visualization

Data monitoring

Measurement Data
Signature Data Extraction – Example

- Low-dimensional gradient filtering
- Windowing based event formation
- Similarity calculation between events w/ DTW
- Similarity matrix for captured events
- Classification of events w/ k-means clustering
- Visual inspection/verification of each cluster
- Signature Library
- microPMU measurement data
- Extracted ‘anomalous’ timestamp
- Extracted ‘anomalous’ events
- k-means clustering
Sample Signature Data

- Blown fuses
- Line recloser open (wire down)
- Recloser close
- Line reclosers open/close (vegetation)

Labeling from analysis of outage records after clustering events
Progress to date

• Collected, reviewed, processed and labeled over 2,600 grid event data from public/private providers, for example
  – DOE/EPRI National Database Repository of Power System Events
  – University of Tennessee Knoxville (UTK) FNET/GridEye Data
  – DOE FOA 1861 Data
    • Eight project teams
    • PMU dataset
      – Covers the three US interconnections
      – Two-year duration
      – Includes event logs

[Signature Summary by Data Provider:

<table>
<thead>
<tr>
<th>No.</th>
<th>Data Provider</th>
<th>Num Signatures</th>
<th>Num Event Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provider 1</td>
<td>295</td>
<td>939</td>
</tr>
<tr>
<td>2</td>
<td>Provider 2</td>
<td>242</td>
<td>319</td>
</tr>
<tr>
<td>3</td>
<td>Provider 3</td>
<td>96</td>
<td>113</td>
</tr>
<tr>
<td>4</td>
<td>Provider 4</td>
<td>143</td>
<td>256</td>
</tr>
<tr>
<td>5</td>
<td>Provider 5</td>
<td>105</td>
<td>377</td>
</tr>
<tr>
<td>6</td>
<td>Provider 6</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>Provider 7</td>
<td>16</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>Provider 8</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Provider 9</td>
<td>1031</td>
<td>2286</td>
</tr>
<tr>
<td>10</td>
<td>Provider 10</td>
<td>663</td>
<td>1097</td>
</tr>
</tbody>
</table>
| Total|               | 2618           | 5467           |]
**Database Metadata & Schema**

- **PostgreSQL + TimescaleDB**

<table>
<thead>
<tr>
<th>Events</th>
<th>Signatures</th>
<th>Waveform</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>eventld</td>
<td>eventld</td>
<td>signatureId</td>
<td>signatureId</td>
</tr>
<tr>
<td>eventType</td>
<td>signatureId</td>
<td>waveType</td>
<td>deviceType</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waveName</td>
<td>duration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waveData</td>
<td>startTime</td>
</tr>
<tr>
<td></td>
<td></td>
<td>waveTimeStamp</td>
<td>stopTime</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>latitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>longitude</td>
</tr>
</tbody>
</table>

**Signature ID: 4**

- **Event Label:** recloser
- **Device Name:** Site0010
- **Sample Rate:** 7680 0 Hz
- **Wave Start Timestamp:** 2006-08-01T00:00:00
- **Wave Duration:** 0.200 sec
- **Event Start Timestamp:** 2006-08-01T00:00:00
- **Event End Timestamp:** 2006-08-01T00:00:00
- **Measurement Types:** Current(A), Voltage(V)
- **Description:** A lightning strike caused a recloser on F_00000026 to operate twice and caused primary wire to fall to the ground.
- **Event Tags:** Equipment:Line wire on ground. Equipment:Interrupting Device::Reclorer. Events:External::Lightning Strike

**Signature ID: 975**

- **Event Label:** Trip
- **Device Name:** Site0010
- **Sample Rate:** 30.0 Hz
- **Wave Start Timestamp:** 2016-10-01T00:00:00
- **Wave Duration:** 239.967 sec
- **Event Tags:** Equipment:Overhead Equipment::Conductor. State::Change State::Trip.
Hierarchical Event Tagging

- Useful for grouping similar types of disturbances
  - Avoids long list of unique disturbance types/conditions
- Flexible & expandable
  - creating entirely new entry when adding new disturbances

<table>
<thead>
<tr>
<th>Group</th>
<th>Events</th>
<th>Conditions</th>
<th>Equipment</th>
<th>State</th>
<th>Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Fault</td>
<td>Weather</td>
<td>Transformers</td>
<td>Steady-State</td>
<td>Single-phase</td>
</tr>
<tr>
<td>• Sub-Class</td>
<td>Momentary, Persistent</td>
<td>Lightning storm, High winds</td>
<td>Voltage regulator, Load tap changer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>Equipment</td>
<td>Equipment conditions</td>
<td>Loads</td>
<td>Change State</td>
<td>Multi-phase</td>
</tr>
<tr>
<td>• Sub-Class</td>
<td>Arcing</td>
<td>Live wire on ground</td>
<td>Inductive, Capacitive</td>
<td></td>
<td>• Two-phase, Three-phase</td>
</tr>
</tbody>
</table>
User Interface

https://gsl.ornl.gov/

Welcome to the GRID SIGNATURE LIBRARY

The Grid Signature Library (GSL) initiative at DOE’s Oak Ridge National Laboratory (ORNL) and Lawrence Livermore National Laboratory (LLNL) is focused on the development of the well-defined, curated, and free-to-access smart GSL to increase the resilience and swift response against malfunctions of grid infrastructure.

- Data visualization & downloading
- Visual analytics of signature data and AI/ML approaches
- Signature matching tool based on AI/ML
- Library API with Shell & Python scripts
- User’s survey
- External data uploading
Signature Matching Tool

• Objective
  – Identify and classify unknown/unlabeled events based on the repository of existing labeled events

• Pre-processing
  – Standardization of measurement data
  – Use variables such as voltage, current and frequency for feature extraction

• Feature extraction
  – Statistical moments: mean, variance, skewness, kurtosis

• Event classification
  – Tested approaches so far
    • Gaussian Naïve Bayes, decision tree, random forest
  – Use unlabeled events in the Library as testing dataset
Basic Classifiers being Tested

Gaussian Naïve Bayes

- Determine class label that maximizes objective function

\[ \hat{y} = \arg \max_y P(y) \prod_{i=1}^{n} P(x_i \mid y) \]

Figure 1. Gaussian Distribution (By M. W. Toews - Own work, based [in concept] on figure by Jeremy Kemp, on 2005-02-09, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=1903871)

Decision Tree

- Find “rules” that separate data in correct class

Figure 2. Example decision tree ("1.10. Decision Trees." Scikit, https://scikit-learn.org/stable/modules/tree.html.)

Random Forest

- Create many “short” trees and vote

Figure 3. Random Forest Algorithm (By Venkata Jagannath - https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=68995764)
Future Works

• Continue to collect real-world power grid event data
  – Perform intensive outreach and engagement with utility professionals
  – Establish and lead community efforts to speed up data collection

• Collaborate with developers of synthetic event dataset
  – Electric Grid Datasets – Texas A&M University
  – pmuBAGE – University of California, Riverside

• Advance AI/ML technologies for grid health monitoring
  – Enable performance testing, benchmarking and comparison

• Facilitate industrial adoption of AI/ML-based approaches
  – Provide an intuitive and visual understanding of AI/ML