Realistic_Synchrophasor_Data_Generation_for_Anomaly_Detection_Using_Cyber-Power_Testbed.pdf

explain the paper in very much detail manner. further clearly explain the dataset in very detailed manner ChatGPT

The paper, titled "Realistic Synchrophasor Data Generation for Anomaly Detection Using Cyber-Power Testbed", presents a sophisticated Hardware-in-the-Loop (HIL) testbed for generating high-fidelity datasets for testing and validating Phasor Measurement Unit (PMU) anomaly detection algorithms. It aims to address the challenges of obtaining realistic, labeled datasets essential for the development of robust anomaly detection tools in modern power systems.

Key Contributions and Details:

1. Objective:

- o The primary goal is to create realistic, labeled datasets for PMU anomaly detection to facilitate training and validation of machine learning algorithms.
- It overcomes challenges such as data privacy, sporadic anomaly occurrences, and the rarity of specific events in real-world grids.

2. The HIL Testbed:

o Components:

- **Real-Time Digital Simulator (RTDS):** Simulates the IEEE 39-bus power system model, including generators, loads, and transmission lines, to replicate real-world grid operations.
- Hardware and Software PMUs: Eight PMUs are strategically placed to ensure grid observability, with two hardware PMUs located at Bus 29 and Bus 39.
- Real-Time Automation Controller (RTAC): Acts as a local Phasor Data Concentrator (PDC) for managing data streams.
- **PingThings Cloud Platform:** Used for real-time data storage and visualization.
- **Network Simulator (ns-3):** Simulates wide-area networking (WAN) to emulate real-world communication scenarios.

3. Data Generation Process:

- o The testbed generates datasets under various operating scenarios, including normal operations, physical faults, and cyber-induced anomalies.
- o **Time Synchronization:** Achieved using the Precision Time Protocol (PTP) to ensure submicrosecond accuracy, critical for reliable PMU data.
- Noise Addition: To simulate real-world conditions, Gaussian noise is added to mimic signal distortions, enhancing the dataset's realism.

4. Generated Dataset Details:

 Features: The dataset includes 14 features such as 3-phase current and voltage magnitudes, corresponding angles, frequency, and the Rate of Change of Frequency (ROCOF).

o Training Data:

• Contains 90 minutes of data sampled at 30 frames per second, resulting in approximately 160,000 data points.

Testing Data:

- Two sets, each with 25 minutes of data and containing specific events (e.g., faults, cyberattacks) at various system locations.
- Approximately 44,000 data points per test set.

5. Event Descriptions:

Training Dataset Events:

• Examples include triple line-to-ground faults, load changes, and concurrent cyber-physical events (e.g., data drops and generation changes).

Testing Dataset Events:

• Events such as load changes (active and reactive), cyber data drops, and combined cyber-physical anomalies (e.g., data drop and line-to-ground fault).

6. Applications:

- The dataset is publicly available on IEEE DataPort and was utilized for an anomaly detection competition during the International Conference on Smart Grid Synchronized Measurements & Analytics (SGSMA) 2024.
- o The dataset supports both supervised and unsupervised learning, aiding in anomaly classification and localization.

Dataset Description:

The dataset is designed for anomaly detection research and consists of realistic, labeled synchrophasor data. It includes:

1. Features:

- 3-phase voltages (magnitude and angle).
- o 3-phase currents (magnitude and angle).
- o Frequency and ROCOF.

2. Training Data:

- o Collected over 90 minutes, comprising ~160,000 data points.
- Scenarios include physical events (e.g., load changes, line-to-ground faults) and cyber events (e.g., data drops).

3. **Testing Data:**

- o Two separate test sets, each with 25 minutes of data (~44,000 data points).
- Events differ in location and type (e.g., generation changes, concurrent cyber-physical faults).

4. Event Labels:

o Label categories include normal data, fault events, line outages, generation changes, load changes, cyber faults, and combined cyber-physical faults.

5. Noise Integration:

- Noise is added using Signal-to-Noise Ratio (SNR) values to mimic real-world measurements.
- For example, 75 dB noise is added to frequency measurements, and 47 dB noise is added to voltage and current measurements.

Conclusion:

The paper highlights the significance of realistic datasets for anomaly detection in PMU data, emphasizing their role in improving grid reliability and security. By integrating advanced hardware and software tools, the testbed offers a versatile platform for generating high-quality datasets under various scenarios, ensuring robust algorithm development. The dataset and testbed contribute significantly to the advancement of anomaly detection research in smart grids.

40

You said:

IEEE_SGSMA_readme.docx

Document

explain the dataset in detail ChatGPT

The dataset described in the **IEEE_SGSMA_readme** document is a realistic labeled dataset for detecting, classifying, and localizing anomalies in Phasor Measurement Unit (PMU) data. It is structured as follows:

Dataset Overview

- The dataset simulates real-time synchrophasor data collected from an IEEE 39-bus power system.
- It includes labeled data for normal operations and various types of anomalies, specifically crafted for anomaly detection in power systems.
- The primary goal is to develop anomaly detection algorithms capable of detecting events, classifying them, and localizing them spatially within the network.

Data Structure

1. Files Provided:

- o **IEEE 39 Bus Power System.raw**: Contains bus, load, source, and branch data for the IEEE 39-bus power system. This is used as the baseline system configuration.
- o **IEEE 39 Bus Power System with PMUs.png**: A diagram of the network topology showing the locations of PMUs.
- o **Event Timeline & Location.xlsx**: Provides event details, including time and location.
- o **PMUbus Location.txt**: Contains PMU placement details alongside bus data.
- o **BusX_Competition_Data.csv**: The main dataset where "X" represents the bus number where the PMU is placed. It includes PMU measurements and event labels.

2. PMU Measurement Data:

- The dataset includes measurements from 8 PMUs, recorded at a rate of 30 frames/second over approximately 90 minutes.
- Features (14 total for each PMU):
 - Voltage Magnitude (in volts)
 - Current Magnitude (in amperes)
 - Phase Angles (in degrees)
 - Frequency (in hertz)
 - ROCOF (Rate of Change of Frequency, in hertz/second)

Labels (Event Types):

- 0: Normal operation
- 1: Fault
- 2: Line outage
- 3: Generation change/outage
- 4: Load change/drop
- 5: Missing data (communication link failure)
- 6: Missing data with a physical event
- 7: Bad data
- 8: Unknown event
- o Multiple events can occur simultaneously in the training and testing data.

Event Detection Tasks

The dataset requires participants to:

- 1. **Event Detection**: Identify when an event occurs.
- 2. **Event Classification**: Classify events into the provided categories.

3. **Event Localization**: Determine the location of events using PMU data.

Dataset Characteristics

- Normal Data: Includes some outliers to represent real-world conditions.
- **Anomalous Data:** Includes a wide variety of labeled events designed to test anomaly detection algorithms.
- Event Labels and Details: Additional details about event timings and locations are provided in the Event Timeline & Location.xlsx file.

Usage Guidelines

- The dataset is intended for anomaly detection competitions and research.
- Participants can use any programming language or technique, including statistical methods, machine learning, and signal processing, to develop algorithms.

The dataset provides a comprehensive framework for exploring the application of PMU data in identifying anomalies, making it valuable for advancing research in smart grid anomaly detection and localization.

40

You said:

Event Number	Event Type	Event Impcat	Approximated Time
9	Physical Event	Generation change	55th min
10	Physical Event	Load change	65th min
11	Physical Event	Load change	70th min

^{*}Corresponding PMU numbers associated with the BUS number is provided in the PMUbus_ Location file in the README folder

explain the dataset

ChatGPT said:

ChatGPT

The dataset in the file contains information about events in a cyber-power testbed environment, detailing specific occurrences and their impacts. Below is an explanation of the dataset structure and contents:

Dataset Description

The dataset tracks various types of events occurring in a testbed setup, along with their impacts on the system. The data includes:

- 1. **Event Number**: A unique identifier for each event.
- 2. **Event Type**: Specifies the nature of the event, such as:
 - o **Cyber Event**: Events related to data or network issues.
 - o **Physical Event**: Events related to physical components, like faults or line outages.
 - o Cyber & Power Event: Concurrent events involving both cyber and power systems.
- 3. **Event Impact**: Describes the specific system impact caused by the event. For example:
 - o Data drops at specific buses (e.g., "Data drop at BUS 29").
 - o Physical faults like "Triple line to ground fault at BUS 39."
 - o Combined impacts like "Concurrent data drop BUS 29 & generation change at BUS 2."
- 4. **Approximated Time**: Indicates when the event occurred during the test or simulation. For example:
 - o "10th min": Event occurred at the 10-minute mark.
 - o "45th min": Event occurred at the 45-minute mark.
- 5. **Event Location**: The specific bus or component affected by the event. Examples include:
 - o "Bus 29"
 - o "Bus 39"
 - o "Bus 24 and Bus 23" (for events affecting multiple locations).

Dataset Insights

- The dataset provides a timeline of cyber-physical events and their locations, which can be used for analyzing system vulnerabilities, event propagation, and response effectiveness.
- Events are categorized by type and impact, enabling focused analyses, such as studying cyber events' effect on data integrity or physical events' impact on grid stability.

If you'd like further exploration of the dataset or specific analyses, let me know!

40

You said:

PMUbus_ Location.txt Document

explain this

ChatGPT

The dataset provides details about a power system test network, specifically a 39-bus load flow test system. Below is a detailed explanation of the dataset:

Key Parameters:

- 1. MVA Base:
 - o The power base for the system is set to 100 MVA. This is a standard base for power system studies.
- 2. Rated Frequency:
 - The system operates at a frequency of 60 Hz, typical for power grids in regions like the USA.
- 3. System Overview:
 - o The dataset describes a **39-bus load flow test system**, including voltage levels, bus types, and other details.

Bus Data:

Each bus in the system is described by:

1. Bus Number:

o Unique identifiers like 'BUS1', 'BUS2', etc., represent the buses in the system.

2. Bus Voltage in kV:

o Voltage levels are given in kilovolts (kV). Most buses operate at 345 kV, while a few operate at 13.8 kV (low voltage levels).

3. **Bus Type:**

- o Indicates the type of bus, generally classified as:
 - Slack Bus (Type 1): Balances the power in the network.
 - PV Bus (Type 2): Maintains voltage magnitude while adjusting reactive power.
 - **PQ Bus (Type 3):** Maintains constant power injection.

4. Per Unit (p.u.) Voltage:

o The voltage magnitude expressed as a per-unit (normalized) value relative to the base voltage.

5. Voltage Angle (Theta):

o The phase angle of the voltage at the bus in degrees.

6. PMU Location:

 Specifies if a Phasor Measurement Unit (PMU) is installed at the bus. Buses without PMUs are marked as 0.

Notable Points:

1. **PMU Installation:**

- o PMUs provide high-resolution measurements of electrical quantities like voltage and current phasors. In this dataset:
 - PMUs are installed at specific buses like BUS10, BUS19, BUS2, BUS22, etc.
 - These PMUs help monitor and enhance situational awareness in the system.

2. Voltage Levels:

- o High-voltage buses mostly operate at 345 kV.
- o Low-voltage buses operate at 13.8 kV, labeled with suffix x1 (e.g., BUS30x1).

3. Voltage Angles:

- Voltage angles vary across buses, influencing power flow directions. For example:
 - BUS1 has a voltage angle of -8.44°, while BUS39 has -10.05°.

4. Low-Voltage Buses:

o Low-voltage buses (13.8 kV) are numbered with x1. These are likely connected to distribution networks or specific equipment.

Applications of the Dataset:

• Power Flow Analysis:

o This dataset is useful for simulating and analyzing power flow in a transmission network.

• PMU-based Monitoring:

o Buses with PMUs can provide real-time data for anomaly detection and grid optimization.

• Research and Testing:

o It serves as a benchmark for validating load flow solvers, contingency analysis, or other power system studies.

If you'd like further clarification on any specific part of this dataset or its application, feel free to ask!

40

You said:

Bus2 Competition Data

TIME	STAMP	BUS2_VA_ANG	BUS2_VA_MAG	BUS2_VB_ANG
0		-65.45328681	208470.11977881388	174.546673
0.033		-65.82644768	209471.99447675075	174.1737033

TIMESTAMP BUS2_VA_ANG BUS2_VA_MAG BUS2_VB_ANG

 0.066
 -66.19925811
 207862.08226040207 173.8006252

 0.1
 -66.57207846
 207839.95663258212 173.4278874

 0.13300000000162981 -66.9450264
 208289.03649929745 173.0551231

 0.1659999999743886 -67.3176271
 210374.61059717086 172.6822492

explain the dataset

ChatGPT said:

ChatGPT

Dataset Overview

The dataset Bus2_Competition_Data.csv contains 161,379 entries and 16 columns. It provides timeseries data collected from Bus 2 in a power system network. Below is a detailed explanation of the dataset's structure and attributes:

Columns Description

- 1. TIMESTAMP:
 - o Type: float64
 - o Represents the time of data collection in seconds.
 - o Range: 0.000 to 5379.266.
- 2. BUS2 VA ANG, BUS2 VB ANG, BUS2 VC ANG:
 - o Type: float64
 - o Represent the voltage angle (in degrees) for phases A, B, and C, respectively.
 - o Range:
 - BUS2 VA ANG: -179.9989 to 179.9999
 - BUS2 VB ANG: -179.9987 to 179.9927
 - BUS2 VC ANG: -179.9955 to 179.9987.
- 3. BUS2_VA_MAG, BUS2_VB_MAG, BUS2_VC_MAG:
 - o Type: float64
 - o Represent the voltage magnitude (in volts) for phases A, B, and C, respectively.
 - o Range:
 - BUS2 VA MAG: 76832.9936 to 229173.4910
 - BUS2_VB_MAG: 78976.8662 to 229533.7544
 - BUS2 VC MAG: 77560.5932 to 230883.4750.
- 4. BUS2 IA ANG, BUS2 IB ANG, BUS2 IC ANG:
 - o Type: float64
 - o Represent the current angle (in degrees) for phases A, B, and C, respectively.
 - o Range:
 - BUS2 IA ANG: -179.9989 to 179.9991
 - BUS2_IB_ANG: -179.9999 to 179.9997
 - BUS2 IC ANG: -179.9993 to 179.9949.
- 5. BUS2 IA MAG, BUS2 IB MAG, BUS2 IC MAG:
 - o Type: float64
 - o Represent the current magnitude (in amperes) for phases A, B, and C, respectively.
 - o Range:
 - BUS2 IA_MAG: 37.9910 to 2163.5660
 - BUS2_IB_MAG: 39.4700 to 2484.8215
 - BUS2 IC MAG: 36.5904 to 2148.5045.
- 6. BUS2 Freq:

- o Type: float64
- o Represents the system frequency in hertz (Hz).
- o Range: 59.2366 to 60.9419.
- o Typical operating frequency is around 60 Hz.

7. BUS2 ROCOF (Rate of Change of Frequency):

- o Type: float64
- o Represents the rate of change of frequency (Hz/s).
- o Range: -423.3942 to 65.4754.

8. Event:

- o Type: int64
- o Categorical indicator of events or anomalies in the system.
- o Range: 0 (no event) to 7 (specific event categories).

Summary Statistics

• Voltage Angles (va ang, vb ang, vc ang):

Show variation from -179.99° to 179.99°, representing the phase shift in the voltage waveform.

• Voltage Magnitudes (va mag, vb mag, vc mag):

Consistently in the range of ~200,000 v, typical for high-voltage buses in power systems.

• Current Magnitudes (IA MAG, IB MAG, IC MAG):

Show variations up to ~2,500 A, indicating load and fault conditions.

• Frequency (Freq):

Fluctuates around 60 Hz, but small deviations are observed, signaling system dynamics.

• Rate of Change of Frequency (ROCOF):

Captures rapid frequency deviations, which can indicate faults or system disturbances.

- Event Classifications:
 - o 0: No anomaly.
 - 1-7: Various types of anomalies or events (specific events not described in the dataset metadata).

Dataset Usage

• Applications:

This dataset can be used for power system monitoring, anomaly detection, and fault diagnosis.

Challenges:

Includes high-dimensional time-series data with multiple phases and metrics, requiring robust preprocessing and feature engineering for machine learning models.

Let me know if you'd like detailed insights or specific visualizations of the dataset.

40

You said:

Bus5 Competition Data

TIMESTAMP	BUS5_VA_ANG	BUS5_VA_MAG	BUS5_VB_ANG
0.0	-68.01805589	202148.29728589283	171.9819098
0.03299999999580905	6-68.39169511	201108.22855840946	171.6084482
0.0659999999916181	-68.76497095	199379.63625179738	171.2349054
0.09999999999126885	6-69.13824915	201957.965964359	170.8617313

TIMESTAMP BUS5_VA_ANG BUS5_VA_MAG BUS5_VB_ANG

0.1330000000162981 -69.51161513 200076.47490996317 170.4885298 explain the dataset in detail

ChatGPT said:

ChatGPT

The **Bus 5 Dataset** consists of 161,379 rows and 16 columns, representing synchrophasor data related to a specific electrical bus (Bus 5). Below is a detailed description of the dataset:

General Information

Total Rows (Samples): 161,379Total Columns (Features): 16

Column Details

1. TIMESTAMP:

- o Type: float64
- o Represents the time at which the data was recorded, measured in seconds.
- o Ranges from 0 to 5379.266 seconds.

2. BUS5_VA_ANG, BUS5_VB_ANG, BUS5_VC_ANG:

- o Type: float64
- o Represents the angular position (in degrees) of phase A, B, and C voltages, respectively.
- o Ranges from -180 to 180 degrees.

3. BUS5_VA_MAG, BUS5_VB_MAG, BUS5_VC_MAG:

- o Type: float64
- o Represents the magnitude of phase A, B, and C voltages, respectively.
- o Typical values range around 199,700 (in Volts), with variations due to system conditions.

4. BUS5_IA_ANG, BUS5_IB_ANG, BUS5_IC_ANG:

- o Type: float64
- o Represents the angular position (in degrees) of phase A, B, and C currents, respectively.
- o Ranges from -180 to 180 degrees.

5. BUS5_IA_MAG, BUS5_IB_MAG, BUS5_IC_MAG:

- o Type: float64
- o Represents the magnitude of phase A, B, and C currents, respectively.
- o Typical values range around 793 Amperes, with slight variations.

6. **BUS5_Freq**:

- o Type: float64
- o Represents the frequency of the power system at Bus 5 (in Hz).
- o The mean frequency is approximately 59.96 Hz, close to the standard 60 Hz. Small deviations are observed due to system disturbances.

7. BUS5_ROCOF (Rate of Change of Frequency):

- o Type: float64
- o Indicates the rate of change of frequency (in Hz/s) at Bus 5.
- o Displays a wide range, from -55.90 Hz/s to 95.02 Hz/s, reflecting significant system events.

8. Event:

- o Type: int64
- o Encodes the type of event occurring in the system (e.g., normal operation or disturbance).
- o Values range from 0 to 4, where:
 - 0 indicates normal operation.
 - 1-4 represent different types of disturbances or events.

Summary Statistics

- The voltage magnitudes (BUS5_VA_MAG, BUS5_VB_MAG, BUS5_VC_MAG) and current magnitudes (BUS5_IA_MAG, BUS5_IB_MAG, BUS5_IC_MAG) exhibit stable behavior with standard deviations of ~1590 and ~15.8, respectively.
- Angular positions show uniform distributions between -180 and 180 degrees.
- Frequency deviations are minimal, with a standard deviation of 0.024 Hz.
- The Event column indicates that most data points represent normal operation (Event = 0), with some samples corresponding to disturbance events.

Use Case

This dataset is useful for analyzing power system dynamics, anomaly detection, and event classification. The features represent real-world synchrophasor measurements collected at Bus 5 and can be used to monitor the state of the electrical grid or train machine learning models to classify system events.

40

You said:

Bus2 Competition Data

TIMESTAMP BUS2_VA_ANG BUS2_VA_MAG BUS2_VB_ANG

0.1999999999970896 -163.522433	209065.18630629068 76.47756047
0.233000000001746 -163.8912763	209120.7142017963 76.10858577
0.2660000000032596 -164.2604216	208652.14126334005 75.73972211
0.3000000000029104 -164.6294284	208284.6610125854 75.37055709
0.3329999999987194 -164.9983283	207714.29294271275 75.00154286
explain the dataset in detail	

ChatGPT said:

ChatGPT

The **Bus2_Competition_Data.csv** dataset contains 43,472 rows and 15 columns. Below is a detailed explanation of the dataset:

Columns in the Dataset

1. TIMESTAMP:

This column represents the time points (in seconds) when measurements were recorded. The values are continuous and range from 0.0 to 1449.033 seconds, sampled at regular intervals.

2. BUS2 VA ANG, BUS2 VB ANG, BUS2 VC ANG:

- These columns represent the phase angles (in degrees) of voltages VAV_AVA,
 VBV_BVB, and VCV_CVC at Bus 2 for three phases (A, B, C).
- Range: Values vary between approximately -180° and +180°.

3. BUS2 VA MAG, BUS2 VB MAG, BUS2 VC MAG:

- These columns represent the magnitudes of the voltages VAV_AVA, VBV_BVB, and VCV_CVC at Bus 2 in three phases (A, B, C).
- o Mean values are around 208,000208,000208,000 with variations across the range of approximately 190,000190,000190,000 to 217,000217,000217,000.

4. BUS2 IA ANG, BUS2 IB ANG, BUS2 IC ANG:

These columns represent the phase angles (in degrees) of currents IAI_AIA, IBI_BIB, and ICI CIC at Bus 2 for three phases (A, B, C).

o Values also range between approximately -180° and +180°.

5. BUS2_IA_MAG, BUS2_IB_MAG, BUS2_IC_MAG:

- These columns represent the magnitudes of the currents IAI_AIA, IBI_BIB, and ICI_CIC at Bus 2 in three phases (A, B, C).
- Current magnitudes range from around 510510510 to 183918391839.

6. BUS2_FREQ:

 \circ This column measures the frequency (in Hz) of the voltage waveform at Bus 2. The values are centered around 59.9659.9659.96, indicating that the system frequency is slightly below the nominal 60 Hz60 \, \text{Hz}60Hz.

7. BUS2_ROCOF:

- o Rate of Change of Frequency (ROCOF) at Bus 2. This metric indicates how rapidly the frequency changes.
- \circ Values are mostly close to 0, but extreme values can range from -12-12-12 to +21+21+21.

Statistical Summary of Key Features

- Voltage Angles (BUS2_VA_ANG, BUS2_VB_ANG, BUS2_VC_ANG):
 - o The angles have a wide spread (standard deviation ~104°), and their values alternate between negative and positive regions.
- Voltage Magnitudes (BUS2_VA_MAG, BUS2_VB_MAG, BUS2_VC_MAG):
 - Mean voltage magnitudes across the phases are approximately 208,000208,000208,000, indicating consistent operating conditions with minor phase-level variations.
- Current Magnitudes (BUS2_IA_MAG, BUS2_IB_MAG, BUS2_IC_MAG):
 - These columns highlight the load currents at Bus 2. The median values indicate stable current magnitudes in the range of 583–585583-585, but outliers are present.
- Frequency (BUS2_FREQ):
 - The system operates close to 60 Hz, with minor deviations.
- ROCOF (BUS2 ROCOF):
 - While the majority of values are near zero, significant spikes occur, which might indicate disturbances or anomalies in the system.

Sample Data (First 5 Rows)

Interpretation

- 1. **Anomaly Detection**: This dataset appears to be well-suited for anomaly detection in power systems, particularly due to outliers in the ROCOF, voltage, and current values.
- 2. **Time-Series Analysis**: The continuous timestamps provide opportunities for time-series analysis to detect trends, deviations, or disturbances.
- 3. **Phase Relationships**: Differences between the phases (A, B, C) in terms of angle and magnitude may reveal imbalances or faults in the system.
- 4. **Dynamic Behavior**: The presence of ROCOF and frequency deviations suggests system dynamics that can be correlated with load changes or grid disturbances.

Would you like additional insights, such as data visualization or specific analysis recommendations?