

DataSense: A Real-Time Sensor-Based Benchmark Dataset for Attack Analysis in IIoT with Multi-Objective Feature Selection

Amir Firouzi *, Sajjad Dadkhah, Sebin Abraham Maret, Ali A. Ghorbani

processed_files:

The `processed_files` directory contains processed data and extracted features in CSV format. It includes both benign and attack samples across all devices in the testbed, using time windows ranging from 1 to 10 seconds. Since all features are time-window based, the data is split into fixed slices of x seconds, and features are extracted for each slice.

The **processed_files** directory contains three subdirectories:

- **attack_data**
 - Processed and feature-extracted CSV files for attack scenarios.
 - Covers different types of attacks performed on the testbed devices.
 - The Category (dos, ddos, ...) and name of the attack (tcp-flood, syn-flood) is determined in the columns label2, label3 in the csv. Label4 provides full name of the attack scenario (dos_tcp-flood, etc.)
 - Organized by time windows, allowing comparison of attack behavior across slices.
- **benign_data**
 - Processed and feature-extracted CSV files for benign (normal operation) scenarios.
 - Represents baseline behavior of the testbed devices.
 - Also structured by time windows to ensure consistency with attack data.
- **Checksums**
 - Contains checksum files for verifying the integrity of the processed CSV data.
 - Ensures that files have not been corrupted or altered during storage or transfer.

Usage:

To use the data, a selected file from `attack_data` and `benign_data` should be matched by time window before analysis. For example:

```
attack_data/attack_samples_1sec.csv +  
benign_data/benign_samples_1sec.csv
```

These matched files allow direct comparison of attack and benign behaviors under the same time-slice resolution. Additionally, different pairs of time windows can be used depending on the type of analysis, enabling more flexible evaluation across multiple temporal granularities.

For complete list of features and their description please refer to the Datasense paper or docs folder.