Task 10:

Mining Time Interval Temporal Patterns For Anomaly Detection in ICS

PI: Dr. Nir Nissim (BGU)

Our Goals and Approach

- Goal:

  Developing an accurate anomaly detection model for ICS based on multivariate time series data (MTSD).

- Our proposed approach:

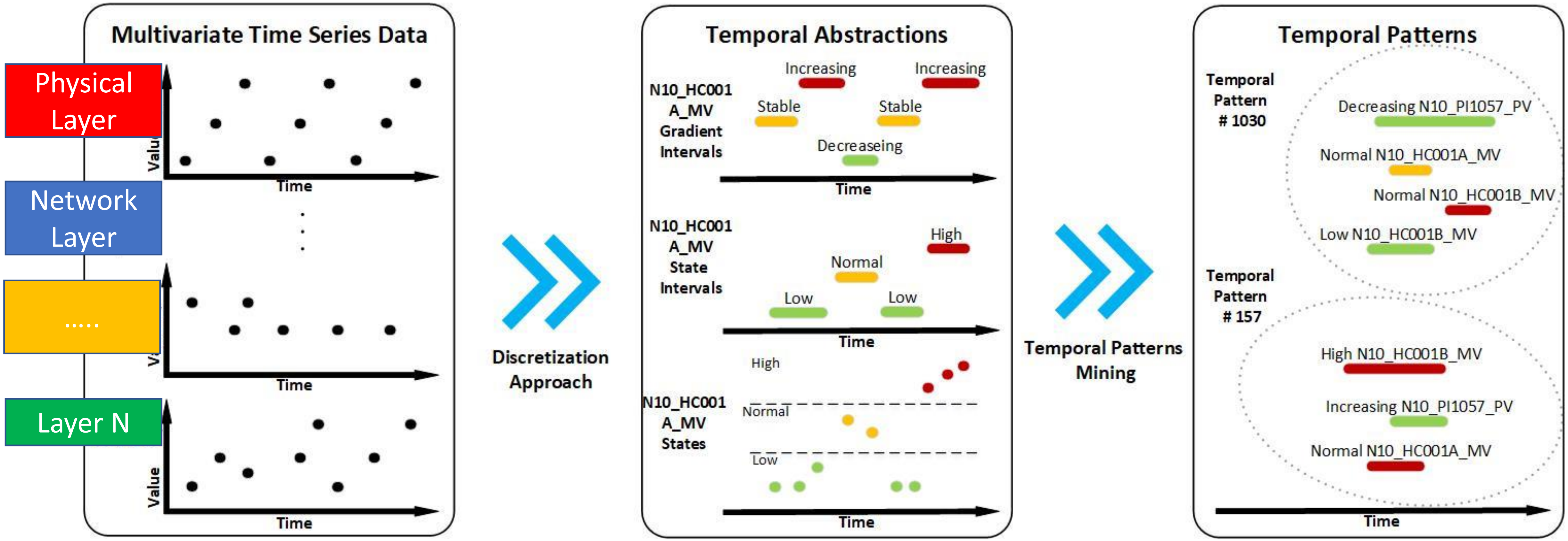
  - Exploiting and Fusing multiple ICS data sources (Physical layer, Network layer etc.)
  - Mining Time Interval temporal patterns that capture the temporal interaction (B. layers and variables)
  - Induce an ML based detection model that well profile normal ICS behavior over time
  - Detect Anomalous behaviors in ICS based on the profiles we have learned

- Current sub goals:

  - Fully understanding the data that we were provided with (Delek, Otorio, DLC)
  - Exploring whether the data is enough for our needs
  - Raising our gaps\inputs regarding the data
  - Receiving further data that meets our needs
  - Designing and Developing our proposed detection model based on the updated data we'll receive

MalwareLab
At the Cyber Security Research Center

Optimal Case: Temporal Patterns Mining from Multiples Sources



Note: vision can be achieved only if all data sources (layers) will be concurrently recorded from same ICS

Current Data: Temporal Patterns Mining from one source separately

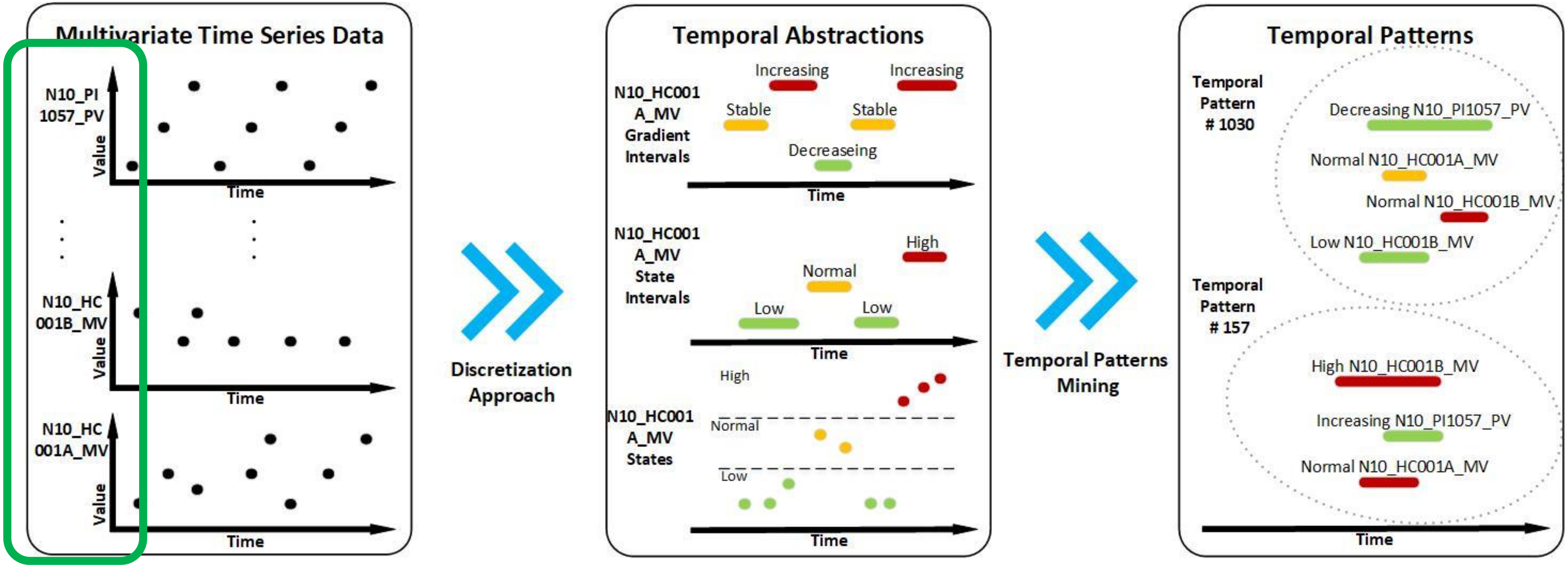
Delek-US PI - Multivariate Time Series Data (MTSD)

- A Recording of a Single ICS entity **(Physical layer)**
- Timestamped data, collected from hundreds of sensors via the PI framework 556 Raw Features (**different** sensor)
- Recording duration ~12 months
- Sampling rate every 30 minutes
- 17521 Timestamped values (in total from all data sources\ sensors)

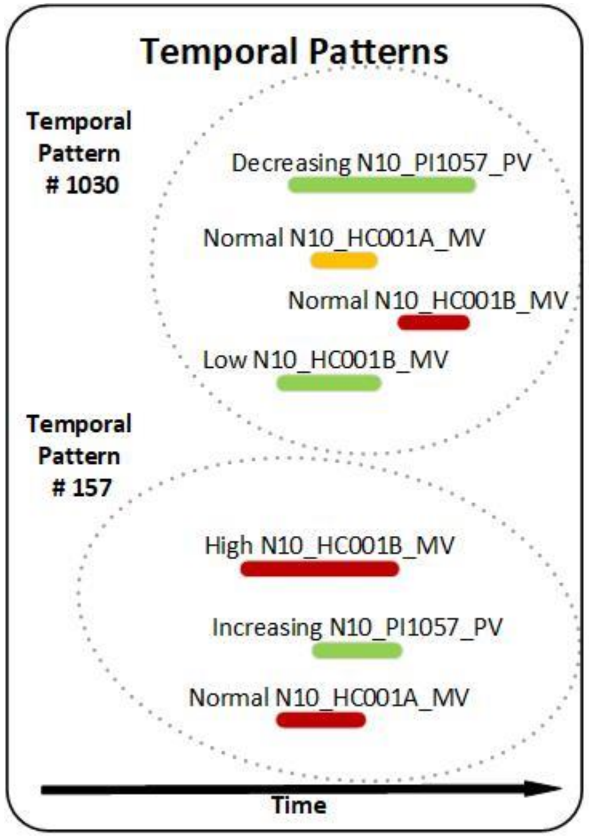
- Our inputs and gaps to be filled:
  - More explanations and descriptions are required regarding the data (Bob?)
  - Higher Sampling rate is required (much lesser than 30 current minutes, e.g. every minute)
  - Data recorded from more **layers are** required to better Profile a Generic Normal Behavior
  - Malicious or Anomalous data should be provided to evaluate the model

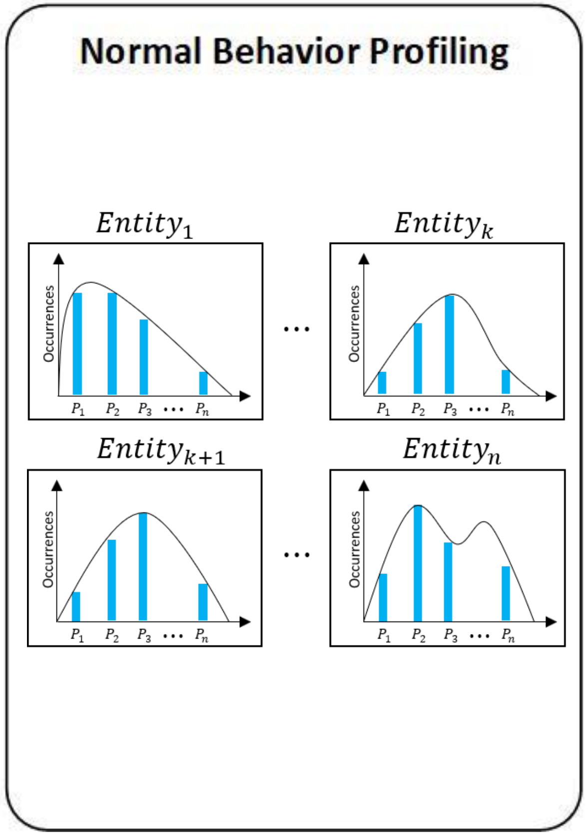
	DateTime	DateTime_Elapsed	DateTime_year	DateTime_month	DateTime_day	DateTime_hour	DateTime_minute	DateTime_second	DateTime_weekday	N10_HC001A_MV	...	N10_FI1053_PV
0	2860.39990	0.000000	2017	10	31	9	35	45	3	58.999939	...	5.422828
1	2860.42065	0.020833	2017	10	31	10	5	45	3	58.999939	...	5.188814
2	2860.44141	0.041667	2017	10	31	10	35	45	3	58.999939	...	4.270872
3	2860.46240	0.062500	2017	10	31	11	5	45	3	58.999939	...	5.009398
4	2860.48315	0.083333	2017	10	31	11	35	45	3	58.999939	...	4.636016

Temporal Patterns Mining – Delek-US MTSD Example



Temporal Patterns For Anomaly Detection – Delek-US MTSD Example





Anomaly Detection

	Test $Entity_1$	Test $Entity_2$	Test $Entity_3$
$Entity_1$	0.5	0.2	0
$Entity_1$	0.1	0.3	0.9
$Entity_1$	0.95	0.8	0.2
$Entity_1$	0.1	0.6	0.5
$Entity_1$	0.77	0.55	0
$Entity_1$	0.34	0.05	0.23

Task 12:

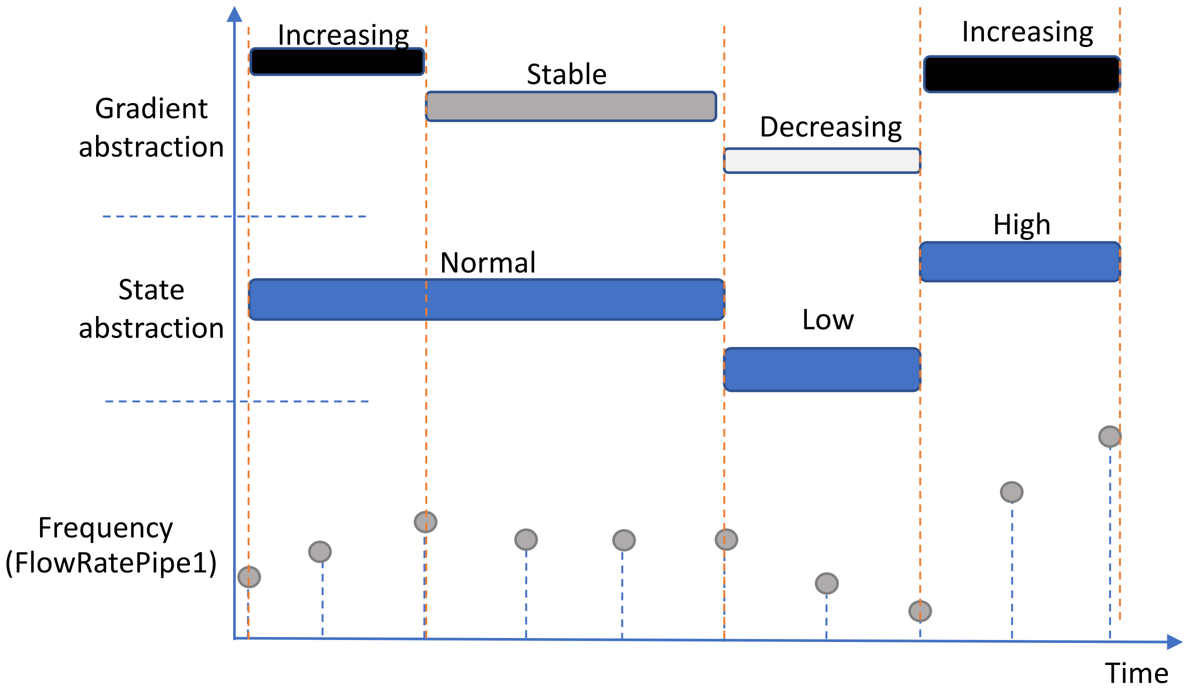
Temporal Explainability of ICS behavior based on Mined Time Interval Temporal Patterns

PI: Dr. Nir Nissim (BGU)

Temporal Patterns Explainability – Developing KBTA for ICS

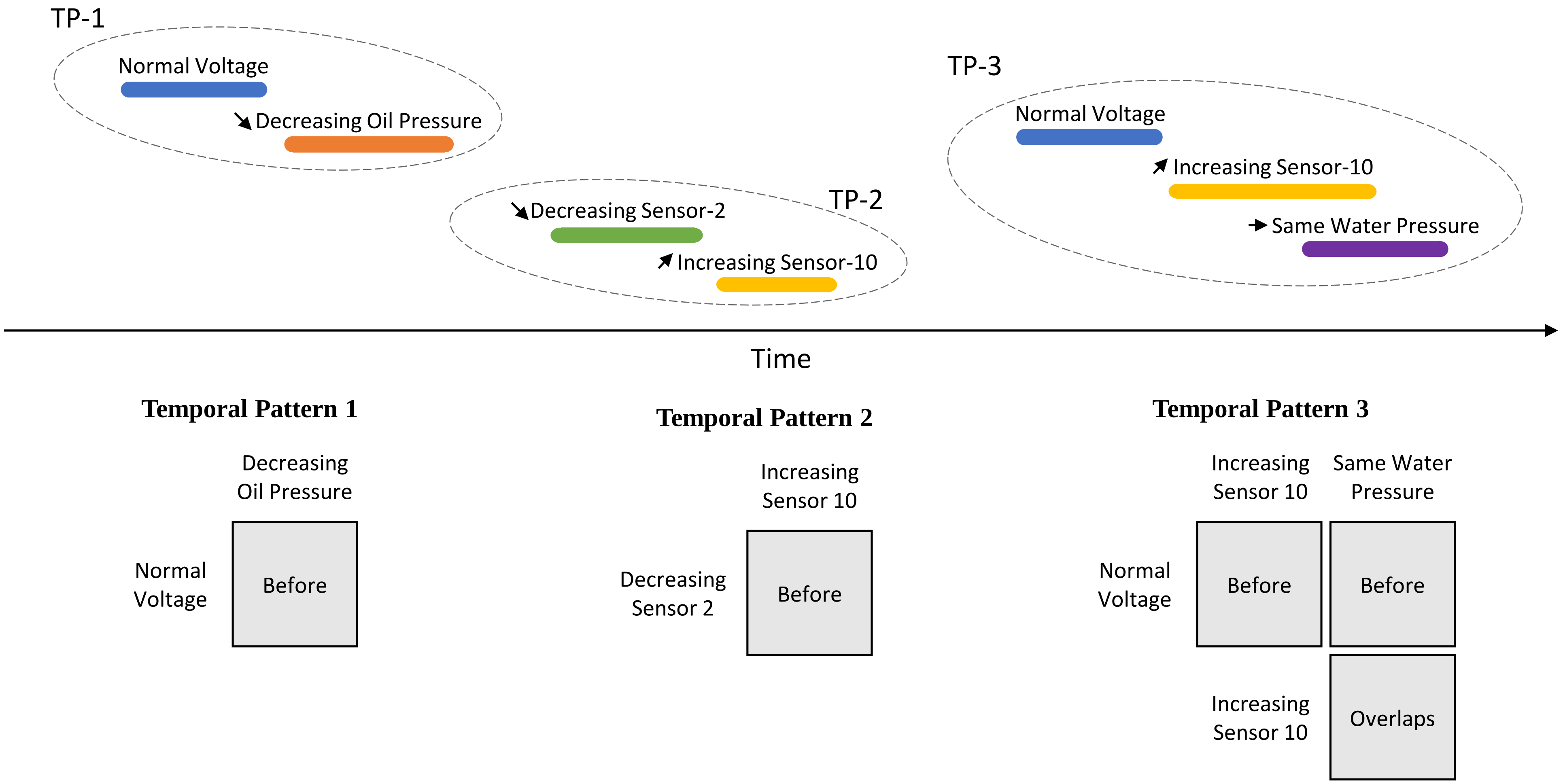
Knowledge Based Temporal Abstraction of an example feature from the WADI Dataset

Feature	Description	Low <	High >	Gradient Trend
FlowRatePipe1	Rate of flow of water into pipe 1 in the testbed	10 m/s	15 m/s	$\lvert\Delta\rvert \geq 1$
FlowRatePipe2	Rate of flow of water into pipe 2 in the testbed	8 m/s	12 m/s	$\lvert\Delta\rvert \geq 1.5$
FlowRatePipe3	Rate of flow of water into pipe 3 in the testbed	20 m/s	30 m/s	$\lvert\Delta\rvert \geq 4$
Boiler1_Temp	Temperature of boiler 1	40° C	110° C	$\lvert\Delta\rvert \geq 0.05$
Tank1_Pressure	Pressure of water tank 1	1e6 Pa	2e6 Pa	$\lvert\Delta\rvert \geq 1e5$



Temporal Patterns Explainability – Developing Exploration & Visualization Module

Highlighting and visualizing interesting TPs across time, one can easily explain temporal behaviors.



MalwareLab
At the Cyber Security Research Center

Experiments and Current Results

for Tasks 10 and 12

Based on WADI Dataset

MalwareLab
At the Cyber Security Research Center

WADI - Multivariate Time Series Data (MTSD)

- The WADI dataset (Ahmed et al. 2017) is a water distribution testbed related data.
- WADI consists of a total of five stages:
  - Three stages controlled by Programmable Logic Controllers (PLCs)
  - Two stages controlled via Remote Terminal Units (RTUs).
- The recorded data consists of:
  - 16 days of sampling (14 normal days ; 2 days containing attack scenarios)
  - 123 measurements (continuous as well as categorial) regarding the testbed:
    - Actuators (valves etc.) related
    - Sensors (pH etc.) related
  - Sampling rates: 60 Hz (each second)
  - 14 different malicious attacks on different parts (e.g. Sensors, Valves, Pumps) of the testbed
  - High class imbalance – 94% of the data is “no-attack” and 6% is “attack”

WADI - Methods and Experimental Design

Data preprocessing:

- Data splitting according to the shortest attack duration (88 seconds)
- 1,980 samples: 120 are related to 1 of 14 attacks , 1880 samples are “no-attack” (Normal)
  → A class balance of 94% Vs. 6%.

Temporal Abstraction & Temporal Patterns mining:

- State abstraction (only) using Equal Frequency Discretization (EFD)
- A vertical support of 50%
- Mining Temporal Patterns of up to size 3 (including)

Machine Learning Algorithms:

- Feature representation using Horizontal Support , Binary
- Feature selection using: Entropy, Gini;
- Selecting different amounts of temporal patterns: 25, 50, 100, 200, 300, 400, 500 and All
- Variety of ML algorithms; RF, SVM (Linear & RBF kernels), KNN, ANN, NB, and LR.

Goal:

- Evaluate the detection capability of our proposed detection method
- Given new **(unlabeled)** time series of ICS data our method should correctly classify to “attack” or “normal”

Experimental Design:

- The learning methods were evaluated using a stratified 5 folds CV
- Classification performance checked correctness of classifying a given new time series an attack or not.
- Classification performance was averaged across the folds and reported on next slide

WADI - Results

Temporal Patterns Mining

- A total of 121,500 time interval temporal patterns have been discovered
  - A total of 105,000 in the "Attack" class of which around 41,000 are exclusive
  - A total of 80,600 in the "No-Attack" class of which around 16,800 are exclusive
  - A total of 63,700 were mutual for both "Attack" and "No-Attacks"

Classification performance (best results are in red)

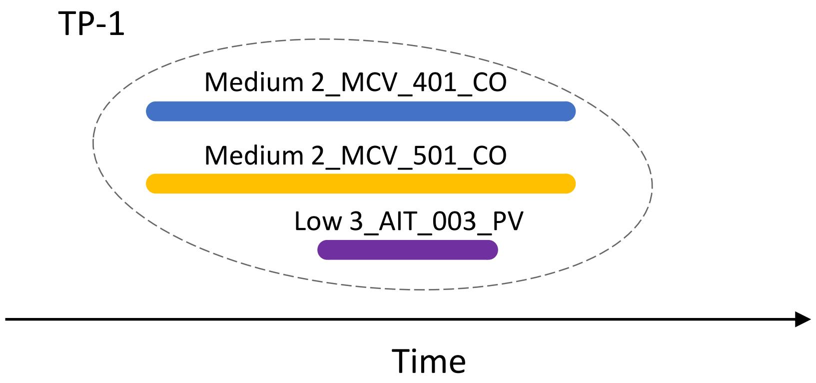
Method	Precision	Recall (TPR)	F1-Score
KNN (k=5), All TPs, Horizontal Support – Best Precision setup	66.26%	74.43%	75.89%
KNN (k=5), Top 100 TPs, Binary, Entropy FS – Best Recall setup	49.21%	87.4%	52.1%
KNN (k=5), All TPs, Horizontal Support – Best F1-Score setup	66.26%	74.43%	75.89%
MAD-GAN (Li et al. 2019) – Best Precision setup	46.98%	24.58	32%
MAD-GAN (Li et al. 2019) – Best Recall setup	6.46%	99%	12%
MAD-GAN (Li et al. 2019) – Best F1-Score setup	41.44%	33.92%	37%

MalwareLab
At the Cyber Security Research Center

WADI - TPs Explainability

Highlighting and visualizing interesting TPs across time, one can easily explain temporal behaviors.

Attack Exclusive TP

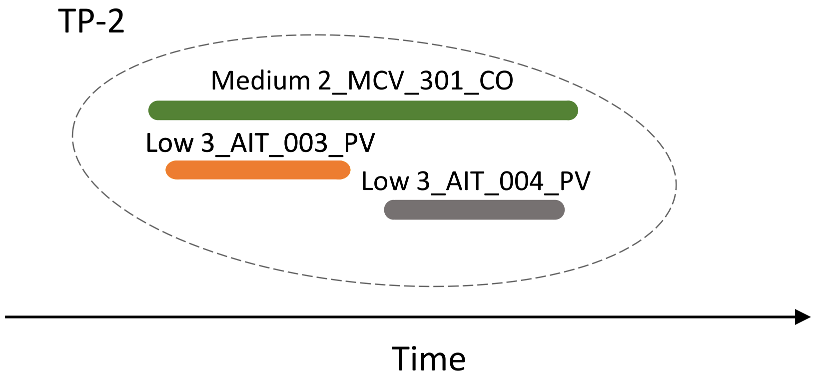


Temporal Pattern 1

	Medium 2_MCV_501_CO	Low 3_AIT_003_PV
Medium 2_MCV_401_CO	Equal	Contain
Medium 2_MCV_501_CO		Contain

MCV = Motorized Consumer Valve
AIT = Sensor to asses compounds in the water

No-Attack Exclusive TP



Temporal Pattern 2

	Low 3_AIT_003_PV	Low 3_AIT_004_PV
Medium 2_MCV_301_CO	Contain	Contain
Low 3_AIT_003_PV		Before

Future Work

WADI data collection (Water Distribution):

Data preprocessing:

- Split the data differently and tune for the best split
- Reduce the number of features using feature selection on the raw data

Temporal Abstraction & Temporal Patterns mining:

- Leverage different discretization approaches (EWD, TD4C or other)
- Leverage additional temporal abstractions (states, gradients)

Machine Learning Algorithms:

- Evaluate additional algorithms as well as TPs dedicated ones (TPF, STF-Mine etc.)

Machine Learning Task:

- Anomaly detection – extend our supervised model

SWAT Data set (Secure Water Treatment)