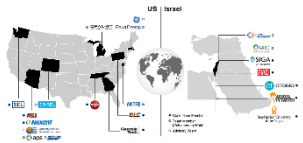
Task 10: Anomaly detection using Pattern Mining

Prof. Robert Moskovitch

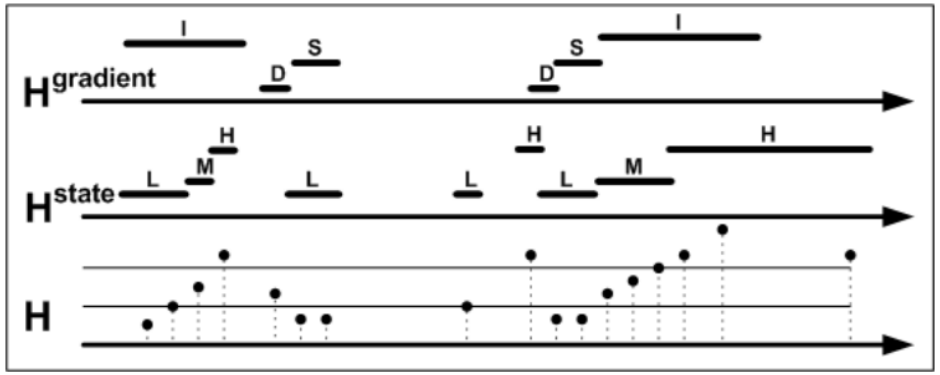
BGU

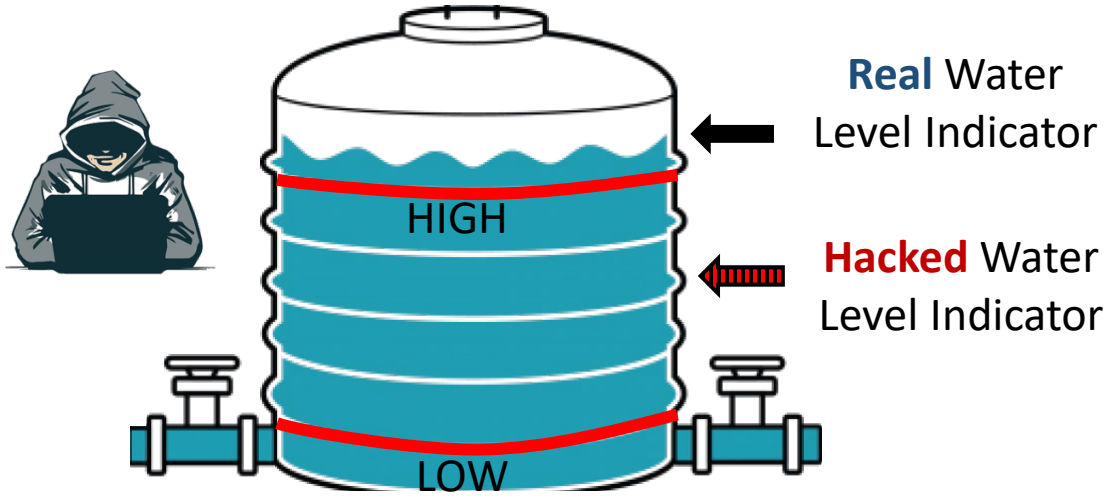
Introduction



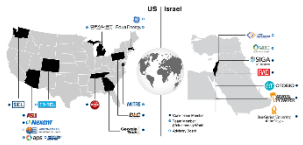
Objective: Develop an anomaly detection model using pattern representation.

- Pattern – A sequence of occurring events over a period of time.





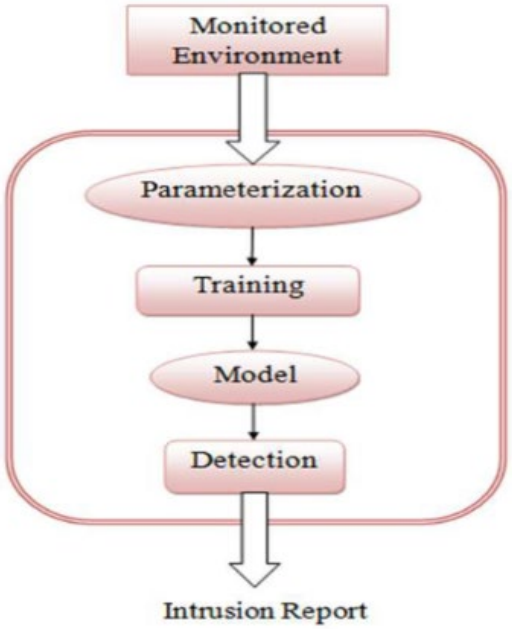
Background



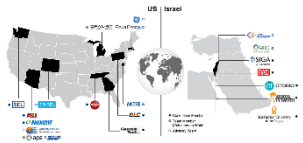
Anomaly detection in industrial systems is challenging due to the complexity of the data.

- A key difference with pattern-based methods is that, unlike deep learning techniques, frequent patterns are easily interpretable.
- Existing pattern-based anomaly detection methods each differ in how they define patterns, support, anomaly scores, and what type of input they are applicable to.

The objective of this research is to propose a new approach for identifying anomalies in data streams using frequent patterns representation.



Sequential Pattern Mining



- Database consists of ordered events, in which each event is 1 time-unit long.

- Apriori property: If a sequence is infrequent, then all its super-sequences must also be infrequent.

Example: < e a > is infrequent, then
< e a b > must be infrequent as well.

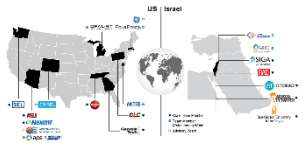
GSP

SPADE

PrefixSpan

SID	Sequences
10	< b a d c >
20	< (bd) a (de) c b >
30	< a d c b >
40	< a d (be) c >
50	< c d b c a b >

Datasets



Secure Water Treatment (SWaT)

Popular dataset, which contains data from a scaled-down water treatment plant with over 20 sensors measuring physical quantities.

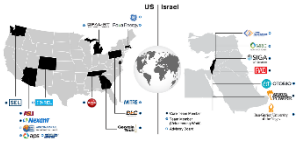
Contains 10 days of normal operational data and 52 types of cyber attacks to evaluate industrial control systems security.

DLC dataset

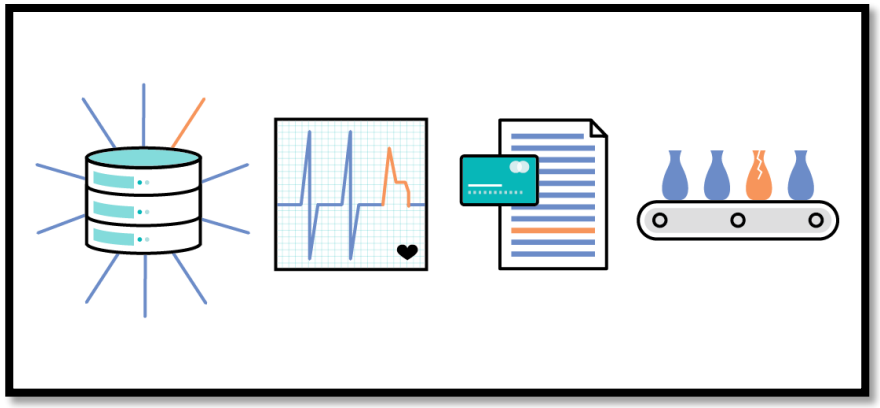
The DLC dataset contains over 51,000 smart meter readings collected in 15-minute intervals over a 29-day period in Pennsylvania, USA.

The data will be used to develop an anomaly detection model to identify unusual energy consumption patterns.

Progress



- Reviewed existing algorithms for pattern-based anomaly detection.
- Collaborated with Task 2 to identify significant and frequent patterns in SWaT operational processes.
- Developing and evaluating our anomaly detection algorithm on SWaT dataset, soon apply on DLC.
- Improving the model using Negative Patterns to identify patterns with higher resolution.
- Experimenting with different configurations for the model, such as window size and number of bins.
- We are now in the process of refining our algorithm and exploring ways to improve its accuracy and efficiency.



Pattern-Based Anomaly Detection

Time Series Data

00:01	00:02	00:03	00:04	00:05	00:06	00:07	00:08	00:09	00:10
0.10	0.15	0.30	0.21	0.32	0.38	0.52	0.59	0.4	0.2

Discretization (a=0.1, b=0.2, ..., e=0.5)

1

a	a	c	b	c	c	e	e	d	b

Sliding Window with Overlap (Window Size=3)

2

a	a	c	b	c	c	e	e	d	b

Tid	Transaction
t1	<a,a,c>
t2	<a,c,b>
t3	<c,b,c>
t4	<b,c,c>
t5	<c,c,e>
t6	<c,e,e>
t7	<e,e,d>
t8	<e,d,b>

Frequent Pattern Mining

3

Pattern	Support
<c>	6
<b>	4
<e>	4
<c,c>	3
<a>	2
<a,c>	2
<c,b>	2
<c,e>	2
<b,c>	2
<e,e>	2
<e,d>	2
<d>	2

Similarity Matrix

4

	<c>	<b>	<e>	<c,c>	<a>	<a,c>	<c,b>	<c,e>	<b,c>	<e,e>	<e,d>	<d>
t1	1	0	0	0	1	1	0	0	0	0	0	0
t2	1	1	0	0	1	1	1	0	0	0	0	0
t3	1	1	0	1	0	0	1	0	1	0	0	0
t4	1	1	0	1	0	0	0	0	1	0	0	0
t5	1	0	1	1	0	0	0	1	0	0	0	0
t6	1	0	1	0	0	0	0	1	0	1	0	0
t7	0	0	1	0	0	0	0	0	0	1	1	1
t8	0	1	1	0	0	0	0	0	0	0	1	1

Anomaly Detector

5

Prediction
Normal
Anomalous
Anomalous
Normal
Normal
Normal
Anomalous
Anomalous

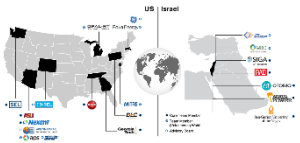
Windows Classification

6

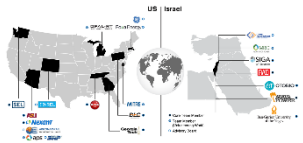
Time Series Data

00:01	00:02	00:03	00:04	00:05	00:06	00:07	00:08	00:09	00:10
0.10	0.15	0.30							
	0.15	0.30	0.21						
		0.30	0.21	0.32					
			0.21	0.32	0.38				
				0.32	0.38	0.52			
					0.38	0.52	0.59		
						0.52	0.59	0.4	
							0.59	0.4	0.2

Future Research Directions



- Explore more integration of Negative Sequential Pattern Mining to improve anomaly detection performance.
- Developing greater similarity measures between transactions and frequent patterns.
- Evaluating algorithm suitability on different domains, such as DLC and other datasets.
- Exploring commercial opportunities to implement our method in real-world settings



Thank You!