

# SSTRESED: Scalable Semantic Trajectory Extraction for Simple Event Detection over Streaming Movement Data

Nikos Giatrakos

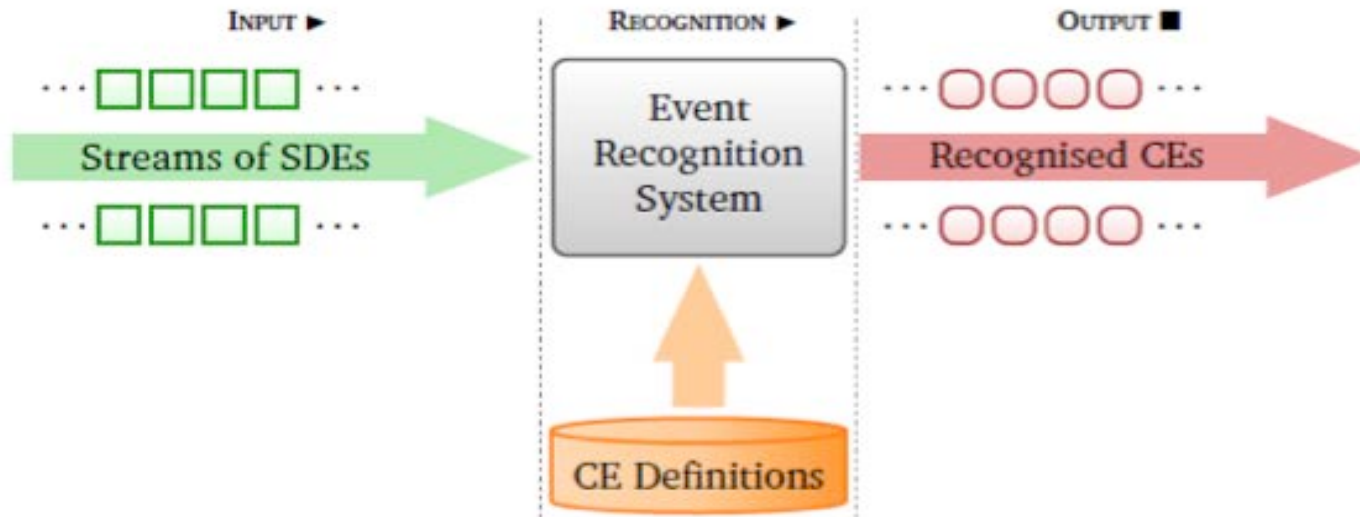
Technical University of Crete

Athena Research Center

*TIME 2023*

*Athens, 26/09/2023*

## Complex Event Detection – The Big Picture

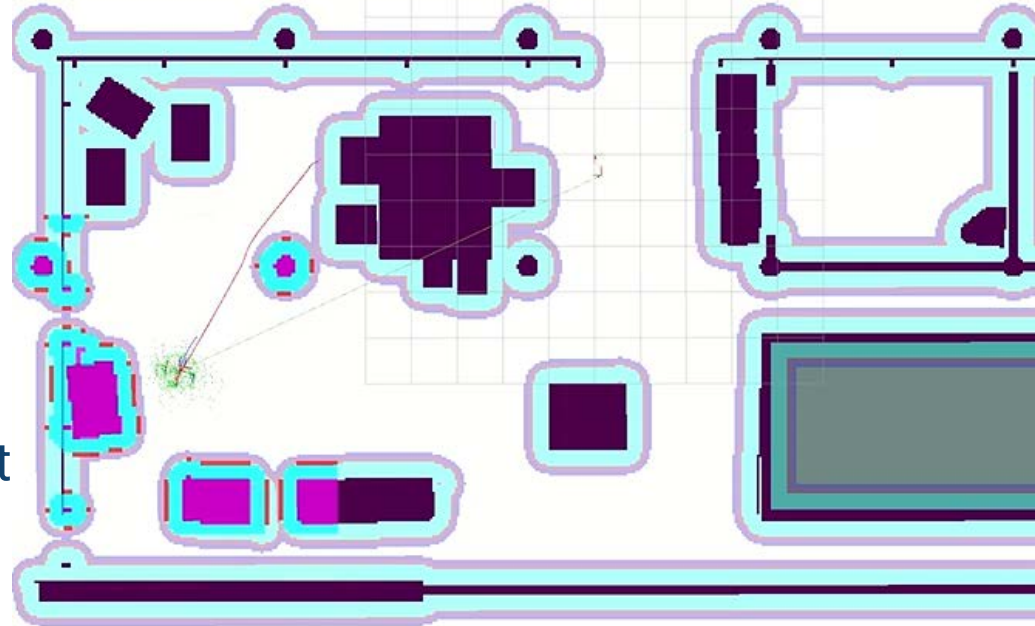


- **Input:** Streams of Simple Derived Events (SDEs)
- **Output:** Complex Events (CEs)
  - Collections of SDEs/CEs satisfying some pattern
  - Patterns defined using a variety of constraints (temporal, spatial, logical,...)

## Motivating Example: Safe Robot Navigation in Smart Factory Scenarios

- SDE Instances:

- Stopped
- Moving
- At Station Si (@Si)
- Rotating
- Goal: Reached Delivery Target
- Blocked: Wall/obstacle
- ...



Credits: DFKI

- Pattern: SEQ(@Si, Moving, Rotating, Stopped, Blocked, Rotating)  
within 5mins

- CE: Delivery Failure

- Aftermath: Delays in production line L

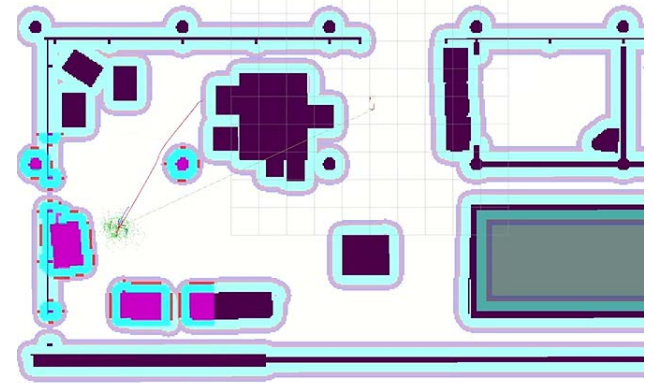
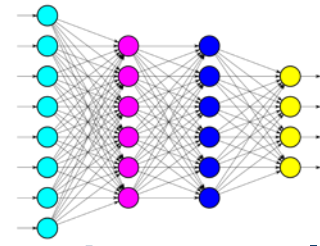


# EVENFLOW Raw Streams: Actual Input

- Raw streams are not SDEs in the real world ☹️

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	...	rot_z	...
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	...	0.9951	...
8.36	753	-3.631	14.922	0.258	0.0979	-0.0019	...	0.9951	...
30.57	3382	-7.473	23.746	0.258	0.8197	-0.0012	...	0.5727	...
38.31	3448	-4.432	23.861	0.258	0.8235	-0.0011	...	0.5671	...
41.15	3704	-7.446	23.866	0.257	0.9951	0.0006	...	0.0977	...
41.12	3705	-7.444	23.867	0.258	0.9952	-0.0001	...	0.0972	...

- Ground Truth - Labeled tuples → Simulations
- Examples of actual SDEs
- ...used to train (NN) models
- ...used to tag unlabeled tuples from the application field



current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	...	Label/SDE
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	...	stopped
8.36	753	-3.631	14.922	0.258	0.0979	-0.0019	0.0001	...	stopped
30.57	3382	-7.473	23.746	0.258	0.8197	-0.0012	0.0015	...	moving to Station2
38.31	3448	-4.432	23.861	0.258	0.8235	-0.0011	0.0013	...	@ Station2
41.15	3704	-7.446	23.866	0.257	0.9951	0.0006	0.0042	....	@ Station2
41.12	3705	-7.444	23.867	0.258	0.9952	-0.0001	0.0032	....	moving to Station3

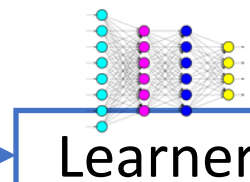
## SSTRESED Version 1.0

### Labeled Time Series - Simulations

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)



PYTORCH



NN Weight Updates  $\begin{bmatrix} w1 \\ w2 \\ \dots \\ wk \end{bmatrix}$



PYTORCH  
Tagger/Classifier

SDEs



Time_from	Time_to	Label
4.25	8.35	moving to Station2
8.35	8.36	stopped at Station2
8.37	30.57	moving to Station4
31.00	38.31	stopped at Station4
39.00	41.15	rotating
42.01	43.12	stopped at Station4

### Unlabeled Time Series - Application Field

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951

— Training Pipeline  
— SDE Detection Pipeline



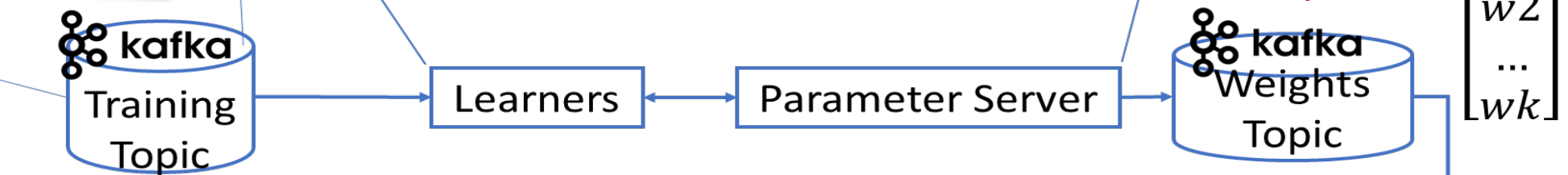
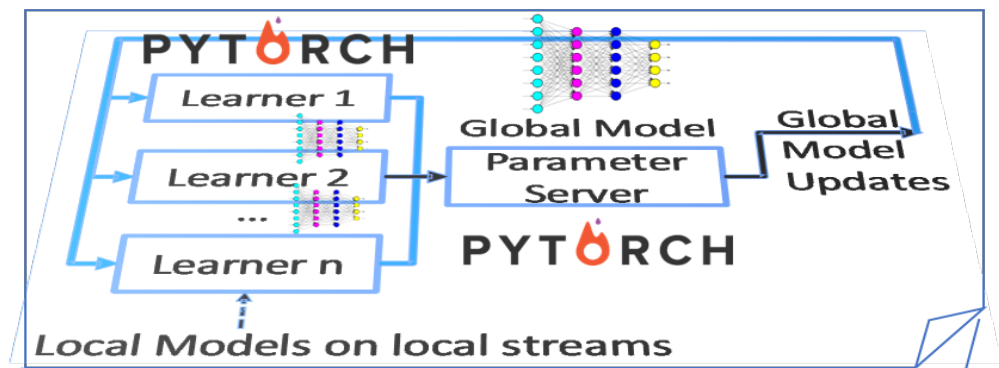
- Support large scale applications – Big Data
  - 1000s of robots producing 100s of MB/min
  - Rapid Continuous Model Training and Online SDE Detection
    - High throughput : number of tuples being processed/tagged per time unit
    - Low latency: amount of time a tuple remains in the Training/SDE detection pipeline
- Solution: Scale-out all kinds of computation
  - Multiple workers for the Training Pipeline
  - Multiple workers for the SDE Detection Pipeline
  - Parallel Processing: Key-based parallelization (Robot ID)

Labeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)
...									

Labeled Time Series @ Partition N

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)

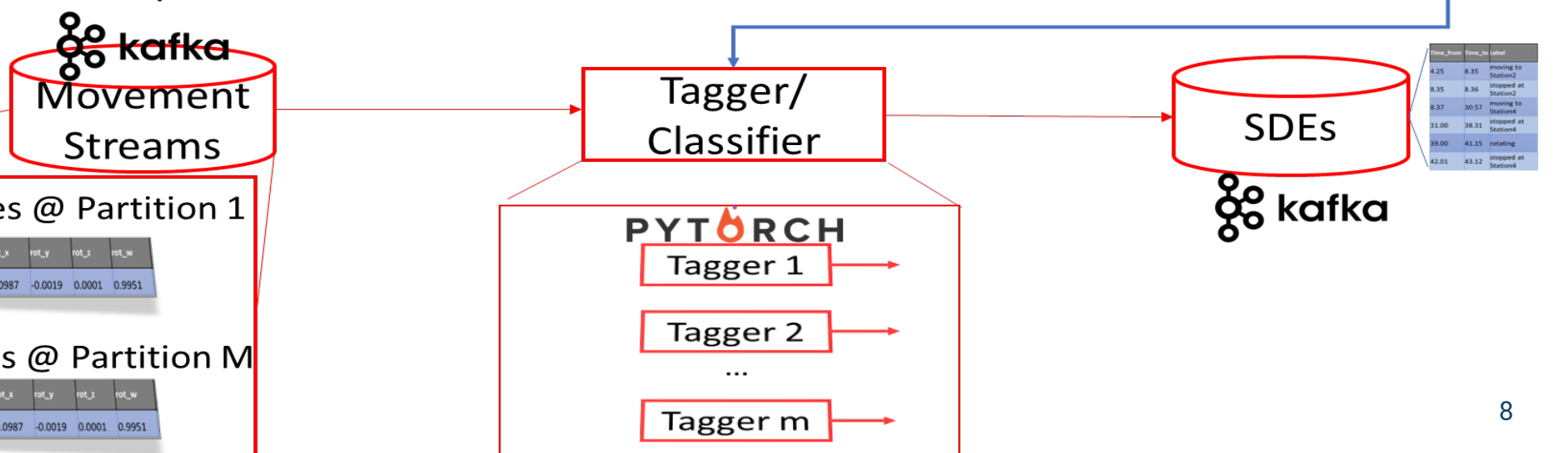


Unlabeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951
...								

Unlabeled Time Series @ Partition M

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951





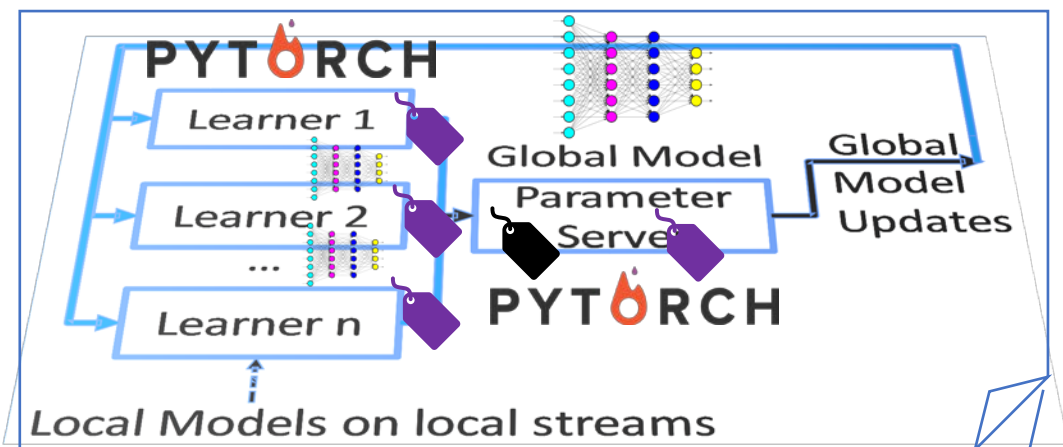
### Labeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)

...

### Labeled Time Series @ Partition N

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)



NN  
Weight  
Updates



$w_1$   
 $w_2$   
...  
 $w_k$



Time_from	Time_to	Label
4.25	8.35	moving to Station2
8.35	8.36	stopped at Station2
8.37	30.57	moving to Station4
31.00	38.31	stopped at Station4
39.00	41.15	rotating
42.01	43.12	stopped at Station4

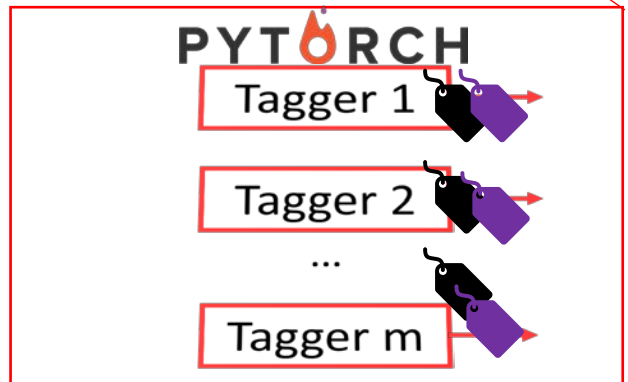
### Unlabeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951

...

### Unlabeled Time Series @ Partition M

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951



- SDEs (not only CEs) are implicitly durative in the Training Pipeline...

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	...	rot_z	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	...	0.9951	stopped
8.36	753	-3.631	14.922	0.258	0.0979	-0.0019	...	0.9951	stopped
10.01	888	...	...	...	...	...	...	...	moving to Station2
30.57	3382	-7.473	23.746	0.258	0.8197	-0.0012	...	0.5727	moving to Station2
35.88	3400	...	...	...	...	...	...	...	moving to Station2
38.31	3448	-4.432	23.861	0.258	0.8235	-0.0011	...	0.5671	@ Station2
41.15	3704	-7.446	23.866	0.257	0.9951	0.0006	...	0.0977	@ Station2
41.12	3705	-7.444	23.867	0.258	0.9952	-0.0001	...	0.0972	moving to Station3

- Required Output of the SDE Detection Pipeline...

Time_From	Time_To	SDE
8.35	8.36	stopped
10.01	35.88	moving to Station2
38.31	41.15	@ Station2
41.12	45.85	moving to Station3
...	...	...

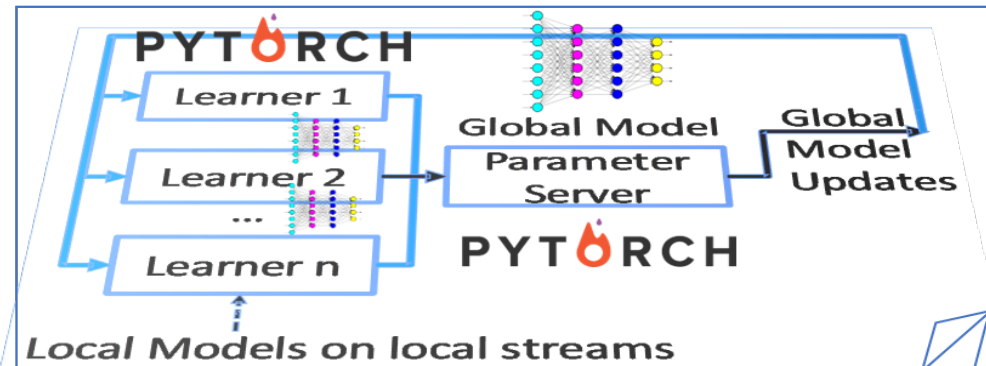
### Labeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)

...

### Labeled Time Series @ Partition N

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w	Label
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951	stopped (unknown)

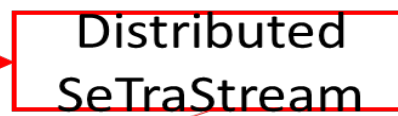


Learners

Parameter Server



NN Weight Updates

$$\begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{bmatrix}$$


Time_From	Time_To	Label
4.25	8.35	moving to Station2
8.35	8.36	stopped at Station2
8.37	30.57	moving to Station6
31.00	38.31	stopped at Station4
39.00	41.15	rotating
42.01	43.12	stopped at Station4

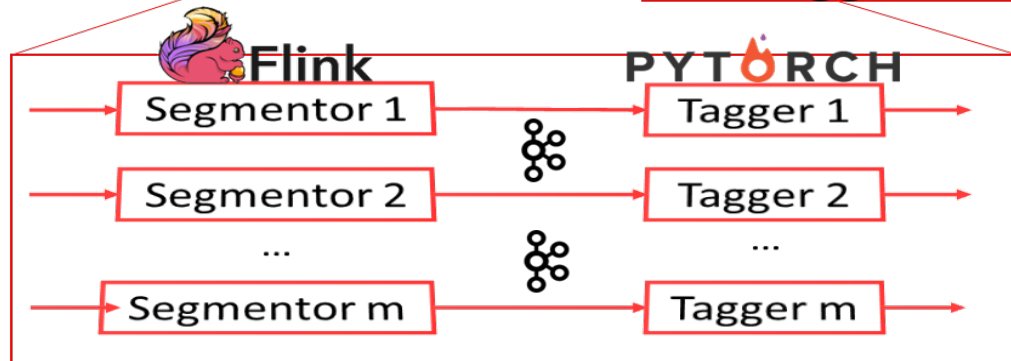
### Unlabeled Time Series @ Partition 1

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951

...

### Unlabeled Time Series @ Partition M

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	rot_z	rot_w
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	0.0001	0.9951



# Movement Feature Vectors (MFVs) Micro-batches

Raw/Original MFVs

current_time	current_step	pos_x	pos_y	pos_z	rot_x	rot_y	...	rot_z	...
8.35	752	-3.626	14.921	0.258	0.0987	-0.0019	...	0.9951	...
8.36	753	-3.631	14.922	0.258	0.0979	-0.0019	...	0.9951	...
30.57	3382	-7.473	23.746	0.258	0.8197	-0.0012	...	0.5727	...
38.31	3448	-4.432	23.861	0.258	0.8235	-0.0011	...	0.5671	...
41.15	3704	-7.446	23.866	0.257	0.9951	0.0006	...	0.0977	...
41.12	3705	-7.444	23.867	0.258	0.9952	-0.0001	...	0.0972	...



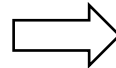
Derived MFVs

speed	direction	acceleration
35 m/s	76°	40 m/s <sup>2</sup>
...	...	...
60 m/s	85°	55 m/s <sup>2</sup>

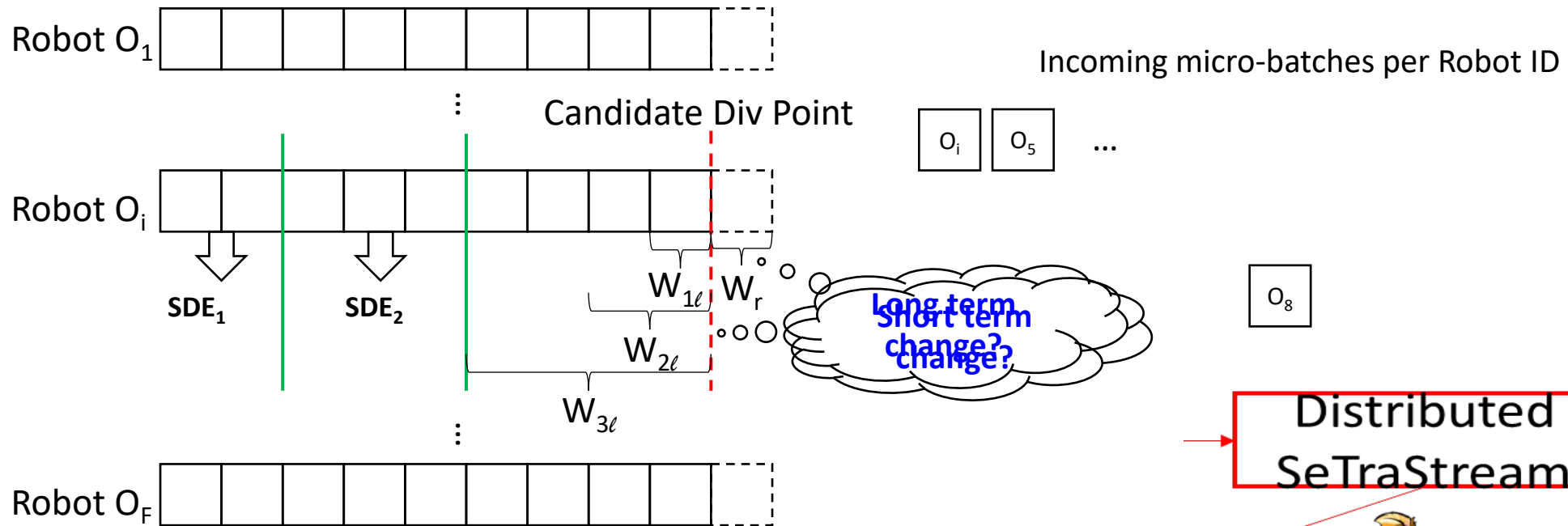


MFVs in micro-batch make up a Matrix

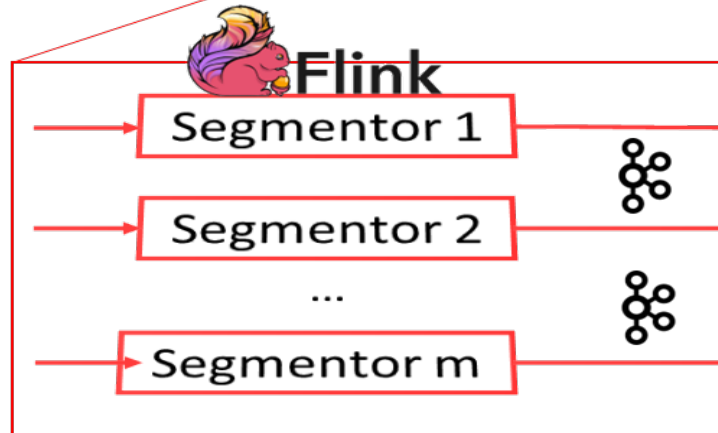
35	...	60
76	...	85
40	...	55
0.1	...	0.05
1	...	3
$\pi/36$	...	$\pi/16$



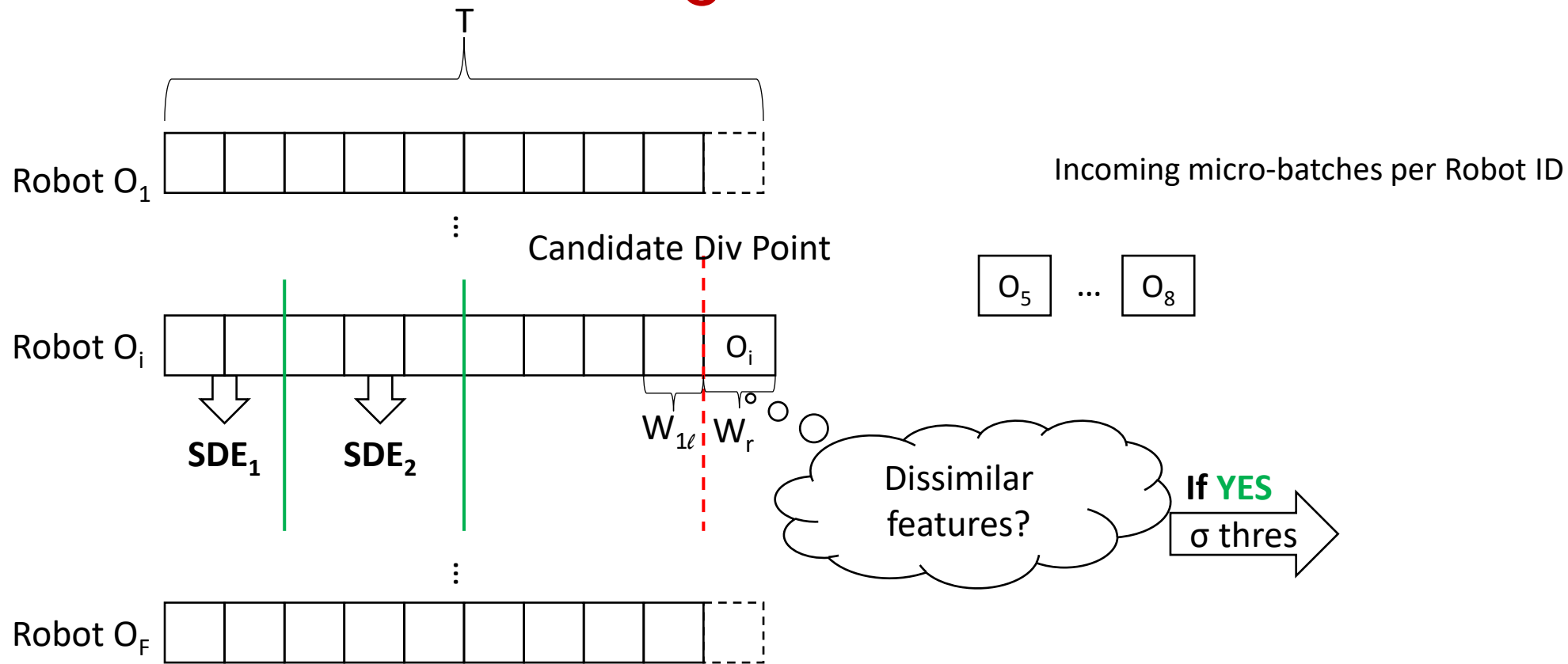
# Distributed SeTraStream - Segmentor View



Distributed SeTraStream



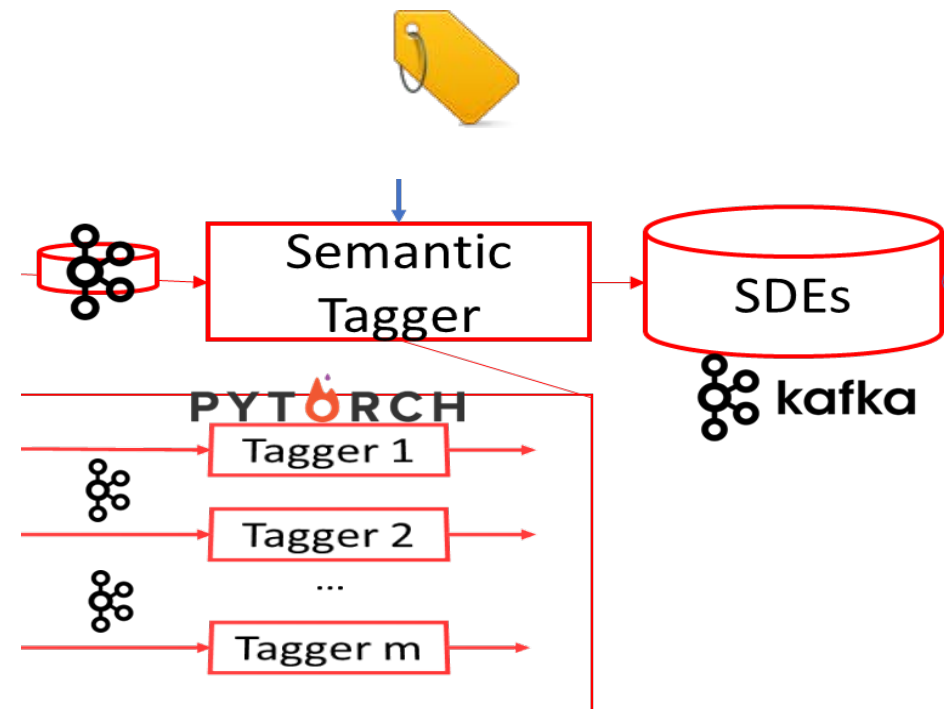
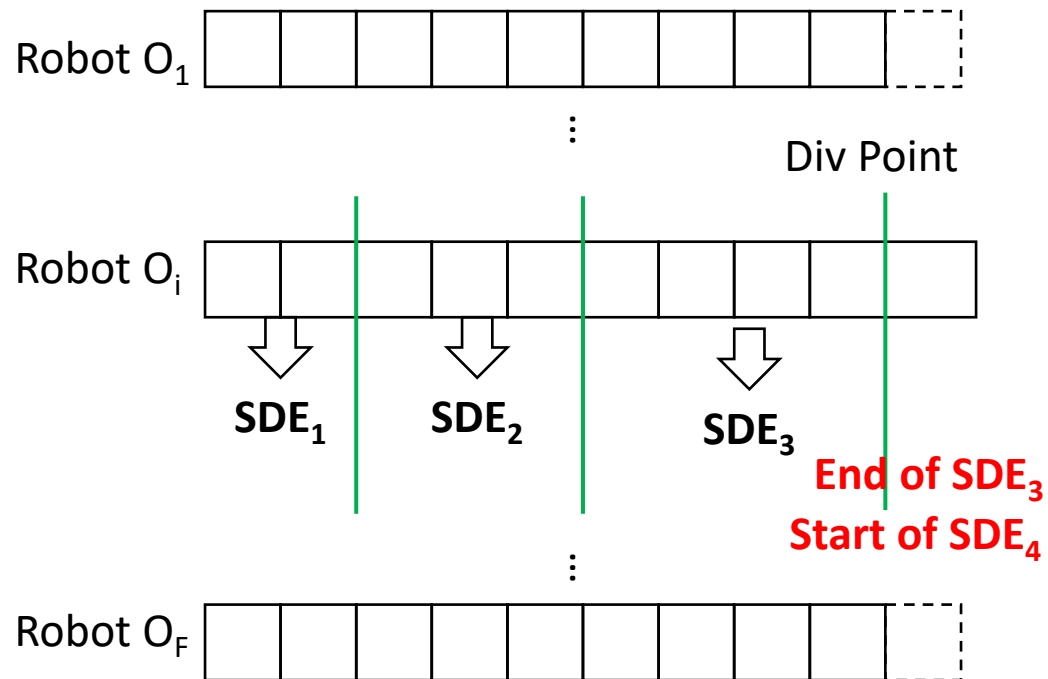
# Distributed SeTraStream – Segmentation Rationale



Similarity Metric  
RV-coefficient ( $W_l, W_r$ )

$$RV(W_l, W_r) = \frac{Tr(W_l W_l' W_r W_r')}{\sqrt{Tr([W_l W_l']^2) Tr([W_r W_r']^2)}}$$

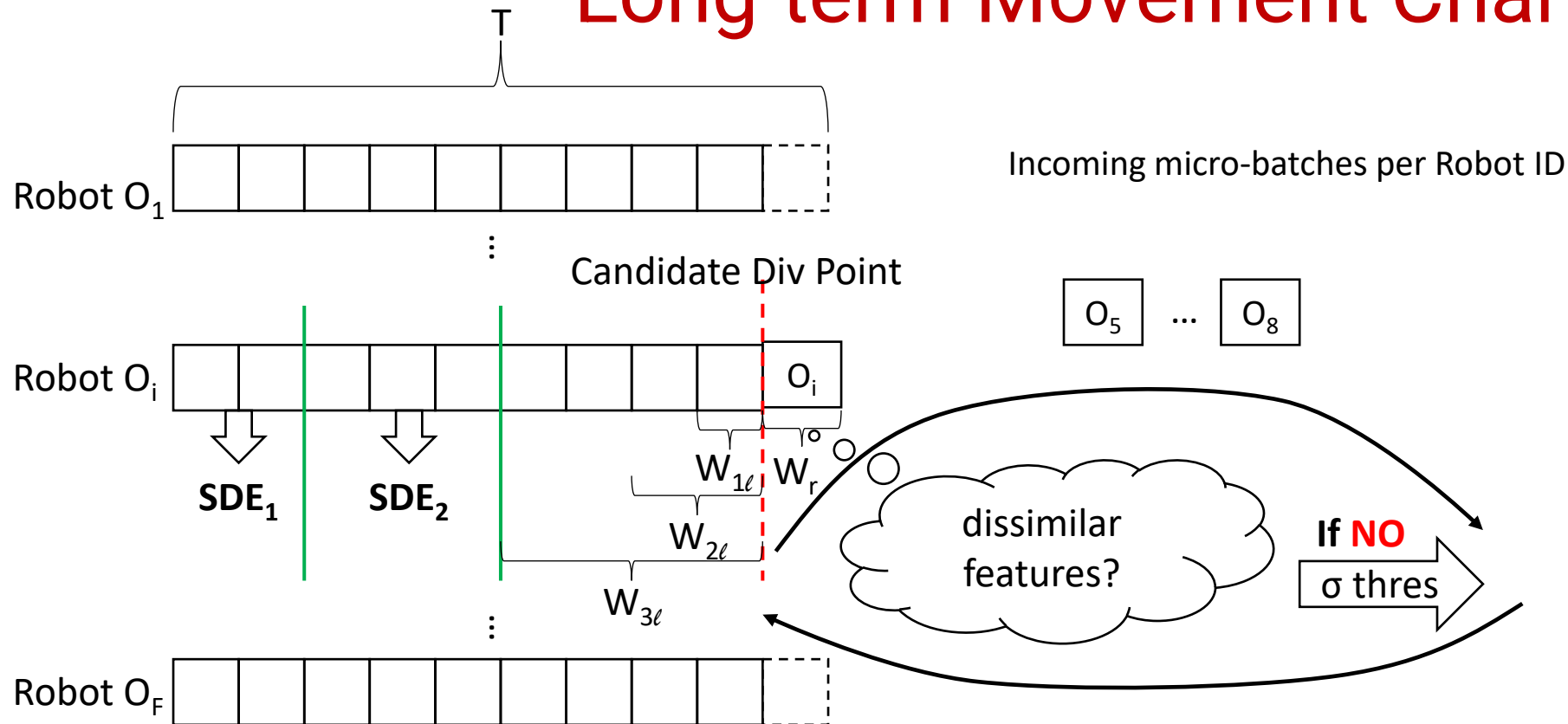
## Distributed SeTraStream - Tagging SDEs



Time_start	Time_end	Label
8.25	8.35	moving to Station2
8.35	8.36	stopped at Station2
8.37	30.57	moving to Station4
31.00	38.31	stopped at Station4
39.00	41.15	rotating
42.01	43.12	stopped at Station4



# Distributed SeTraStream - Long-term Movement Change

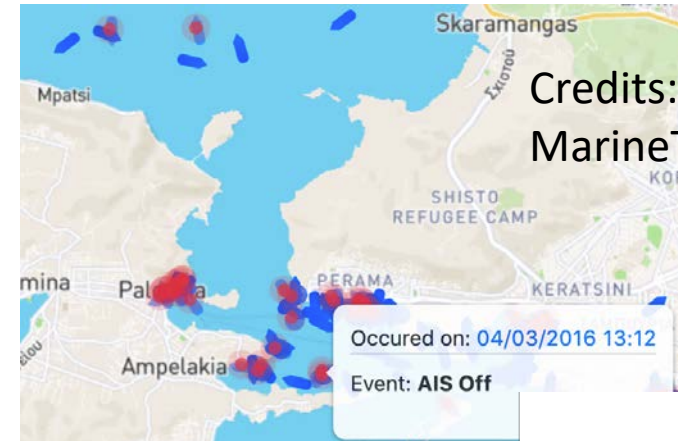


Similarity Metric  
RV-coefficient ( $W_l, W_r$ )

$$RV(W_l, W_r) = \frac{Tr(W_l W_l' W_r W_r')}{\sqrt{Tr([W_l W_l']^2) Tr([W_r W_r']^2)}}$$

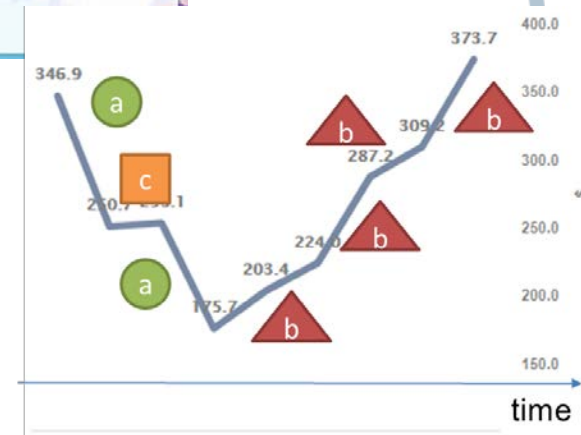
- 1<sup>st</sup> Scalable Streaming Framework for SDE Detection
  - Novel framework & architecture treats SDEs as first-class citizens
  - Distributed processing across state-of-the-art Big Data Technologies
  - 1<sup>st</sup> to combine streaming semantic trajectory extraction and CER

### • More Applications



### • On-going work

- Complete implementation
- Enrich semantics (terrain images, map POIs etc)
- Trajectory compression before segmentation
- Wayeb[VLDBJ'22], EasyFlinkCEP [CIKM'21] Interoperability





# Thank you

Nikos Giatrakos

ngiatrakos@softnet.tuc.gr

ngiatrakos@athenarc.gr