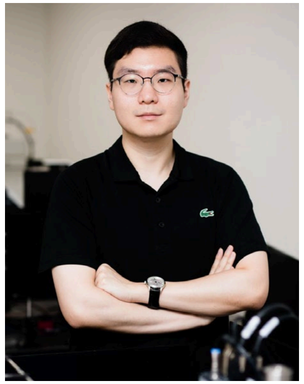
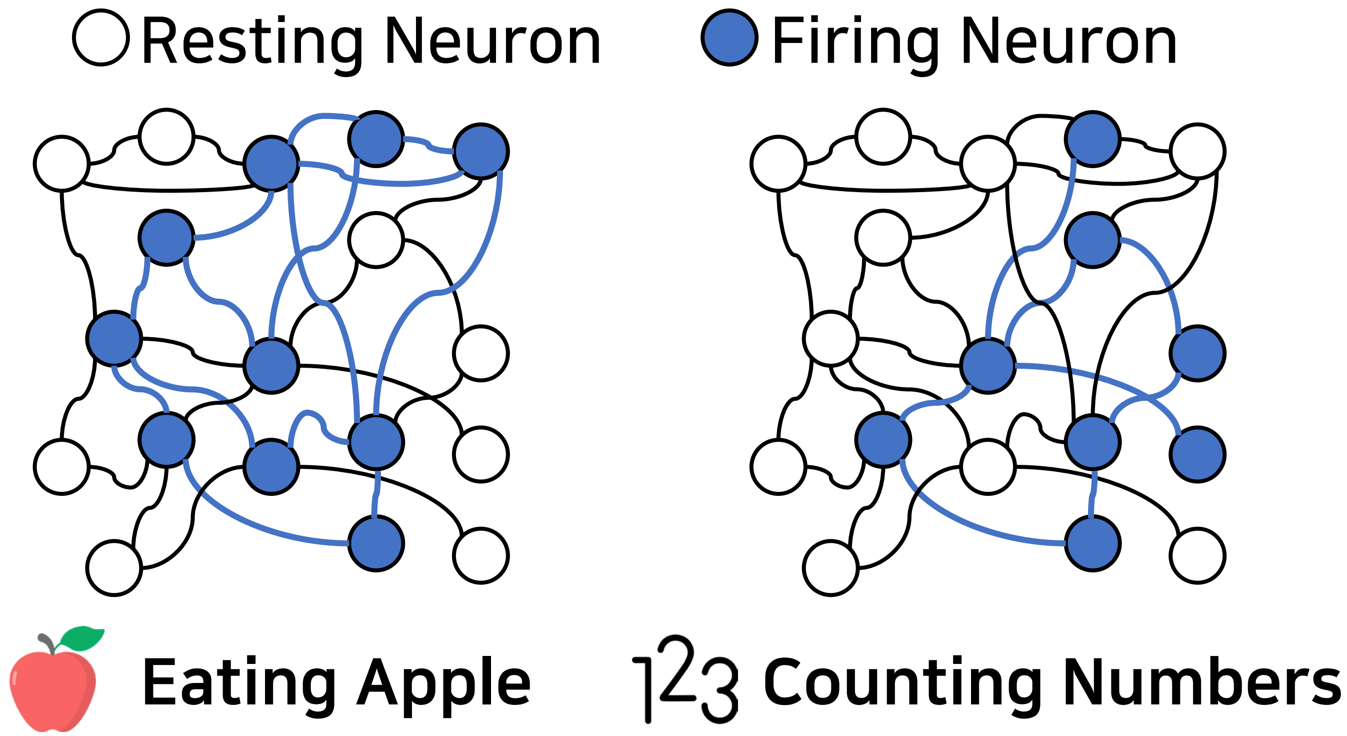
RASP: Robust Mining of Frequent Temporal Sequential Patterns under Temporal Variations


Hyunjin Choo   Minho Eom   Gyuri Kim   Young-Gyu Yoon   Kijung Shin

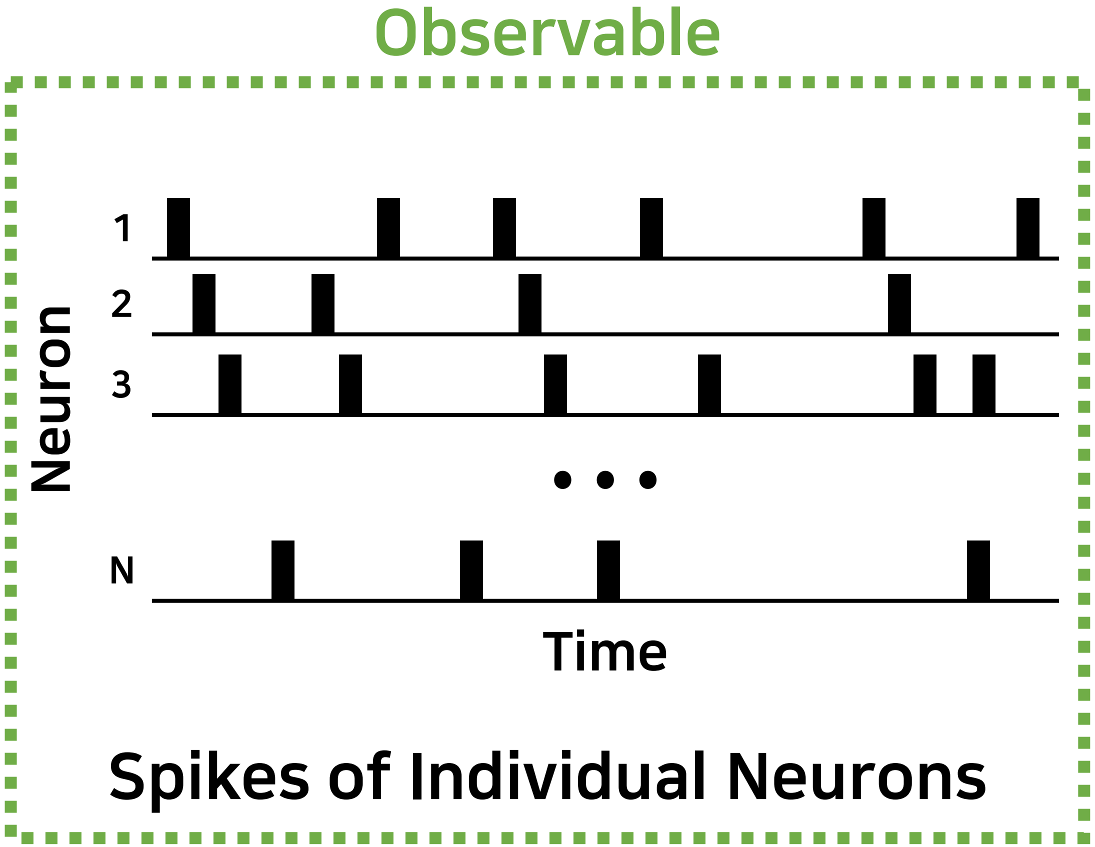
Group Interactions are Everywhere!

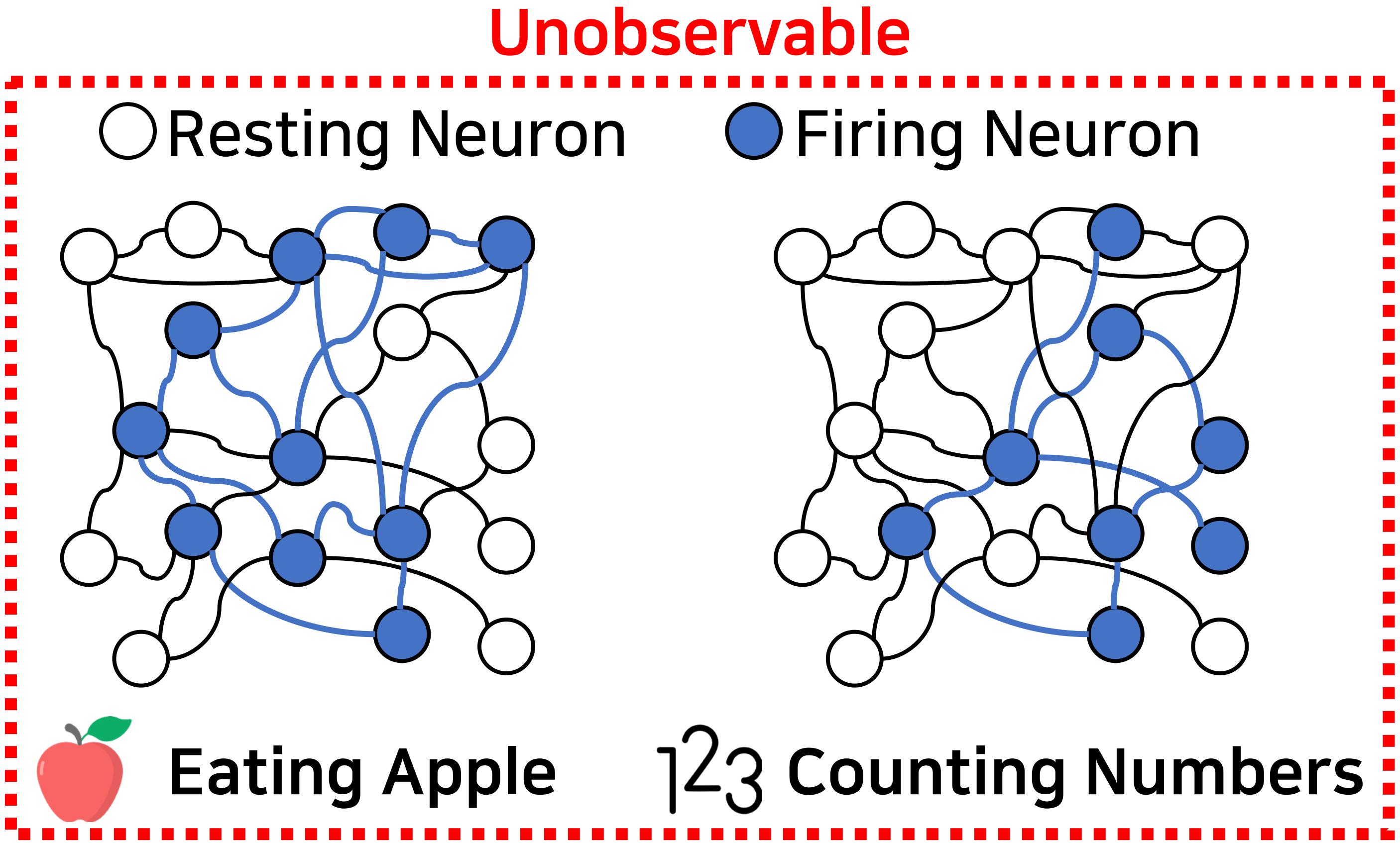
- A group Interaction (GI) is an interaction involving two or more entities
  - E.g. Functional groups of neurons associated with specific tasks



Group Interactions are Often Unobservable

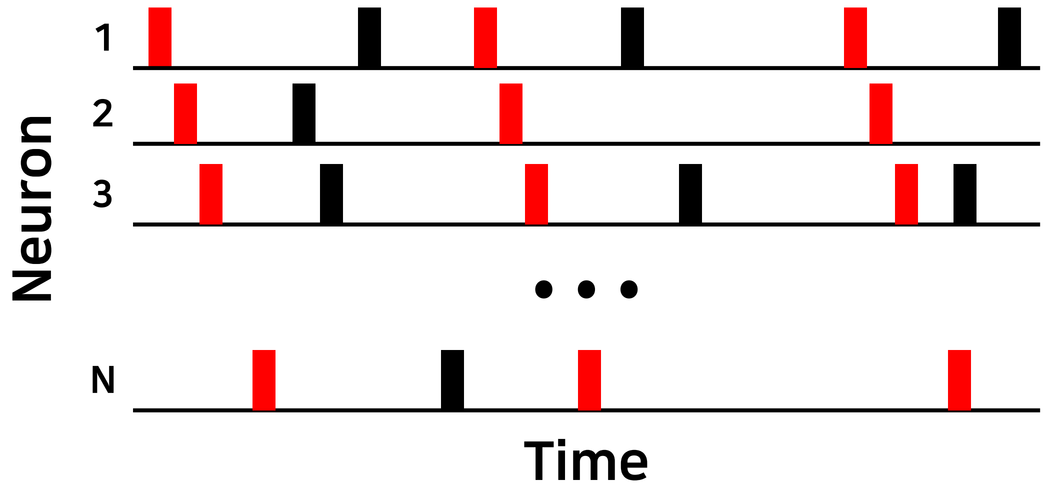
- However, functional groups of neurons associated with specific tasks cannot be directly observed
  - Instead, only the spiking activities of individual neurons can be observed





Characteristics of Group Interactions

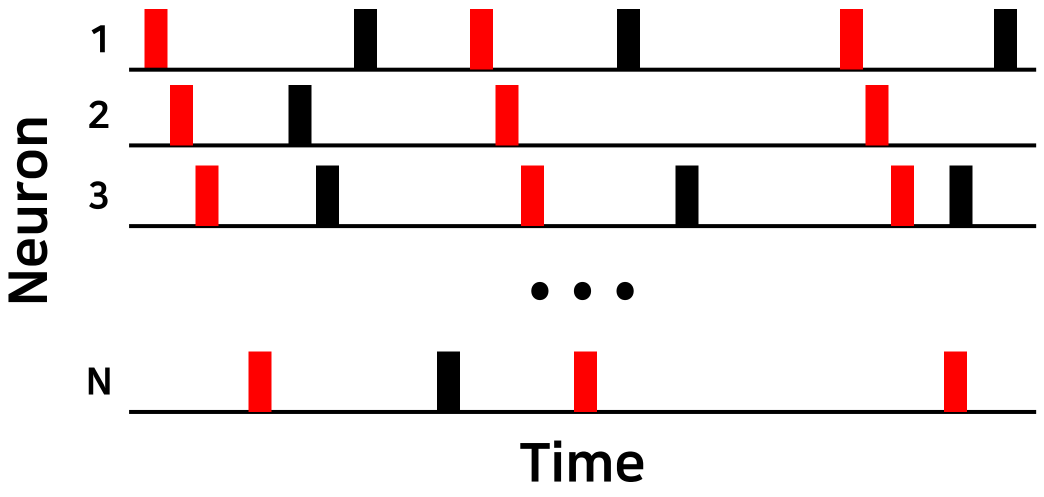
- A group interaction is empirically found as a pattern characterized by correlated events among multiple entities in the observed sequence of events
  - E.g. Functional groups of neurons exhibit temporally correlated spiking activities



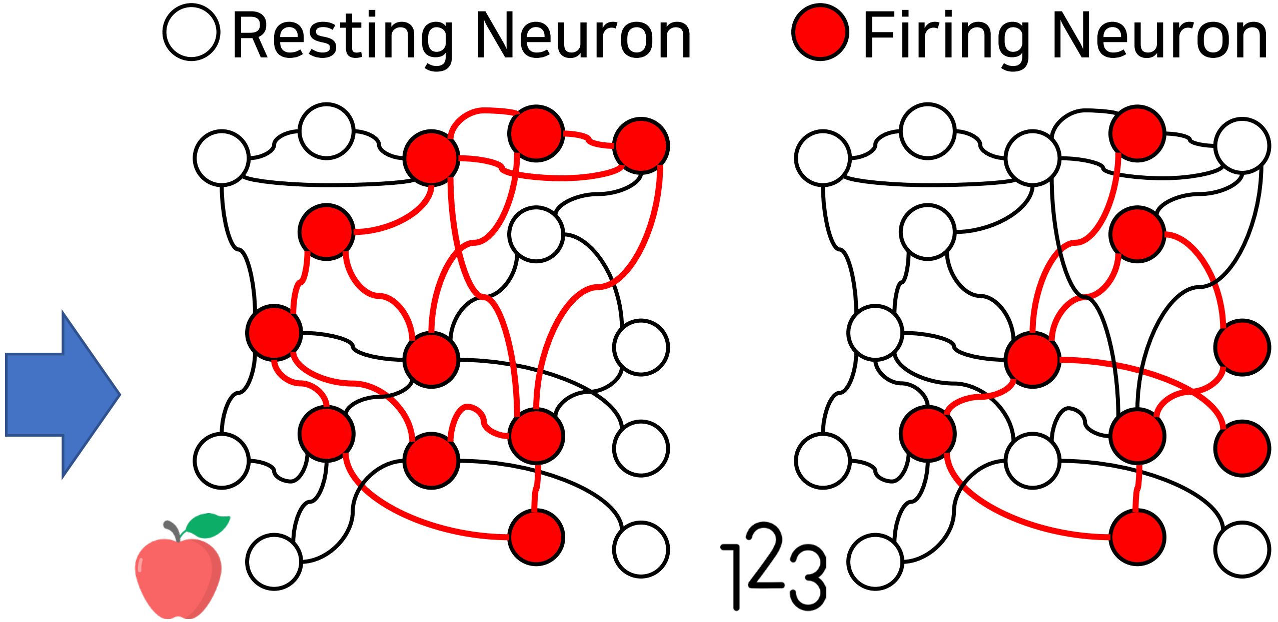
Correlated Spiking Activities

Research Goal

- **Goal**: Given activities of individuals, to identify group interactions (GIs)
  - **Our approach**: To find empirical patterns in the observed events of individual entities
  - E.g. Functional groups of neurons exhibit temporally correlated spiking activities



Correlated Spiking Activities
(Empirical Patterns)

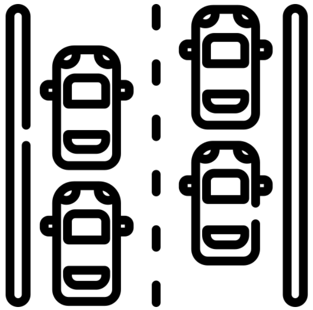


Functional Groups of Neurons
(Group Interactions)

Real-World Applications

- Precipitation at weather stations
  - Target GI: A series of regions experiencing consecutive rainfall events
- Traffic volume at intersections
  - Target GI: A network of roads experiencing successive congestion
- Prices of individual stocks
  - Target GI: A group of stocks exhibiting correlated price movements

Precipitation



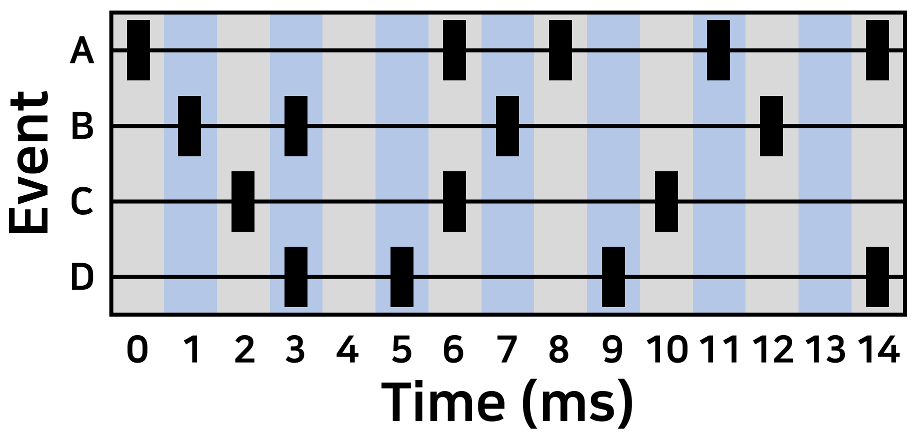
Traffic

Stock

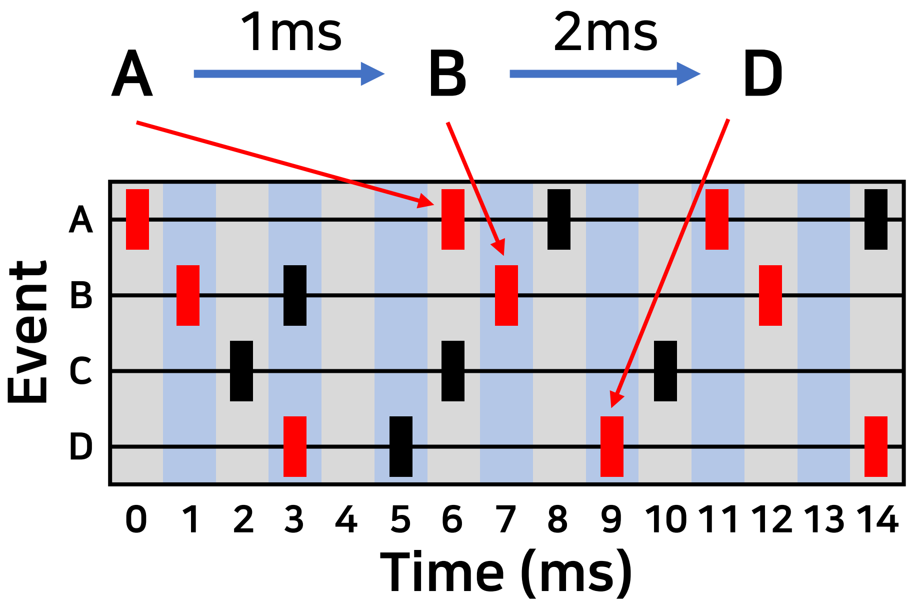
Concepts: Temporal Sequential Pattern

- **Temporal event sequence**: A sequence of event instances ordered by time
- **Temporal sequential pattern (TSP)**: A sequence of events and time gaps
  - Temporally and sequentially correlated events can be represented as a TSP

Temporal event sequence



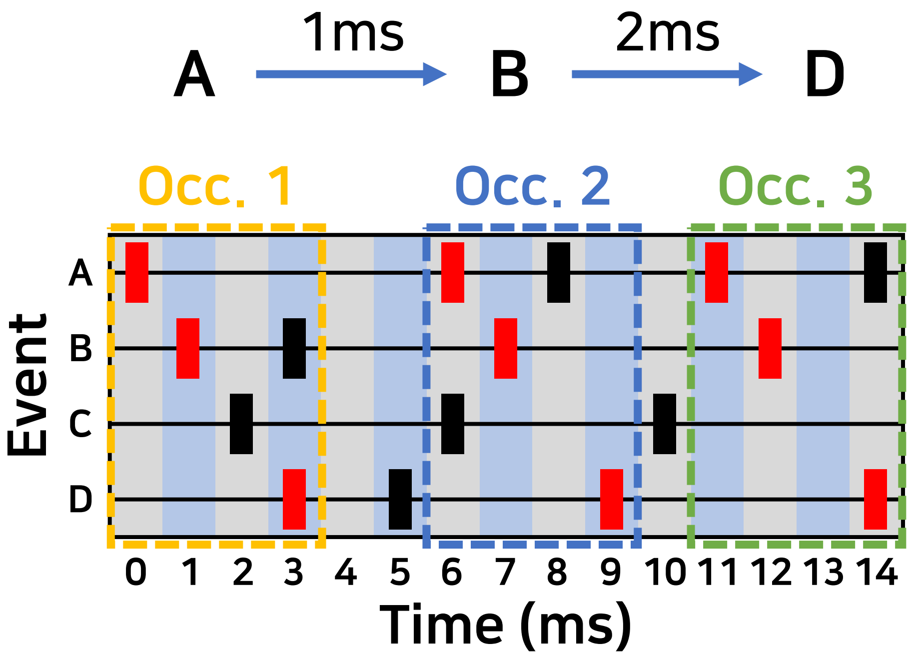
An example of TSP



Occurrence and Support of a TSP

- An **occurrence** (or **instance**) of a TSP: An actual case in which the TSP occurs
- **Support** of a TSP: Number of occurrences of a TSP

An example of occurrence and support of a TSP



Leverage: Significance Measure of a TSP

- **Leverage**: Difference between a TSP's actual and expected support
  - Expected support is computed under the assumption of independent event occurrences

$$leverage(\alpha) = support(\alpha) - support_{exp}(\alpha)$$

Problem Formulation

- **Given**: A temporal event sequence
- **Find**: Group interactions

Problem Formulation

- **Given**: A temporal event sequence
- **Find**: Group interactions

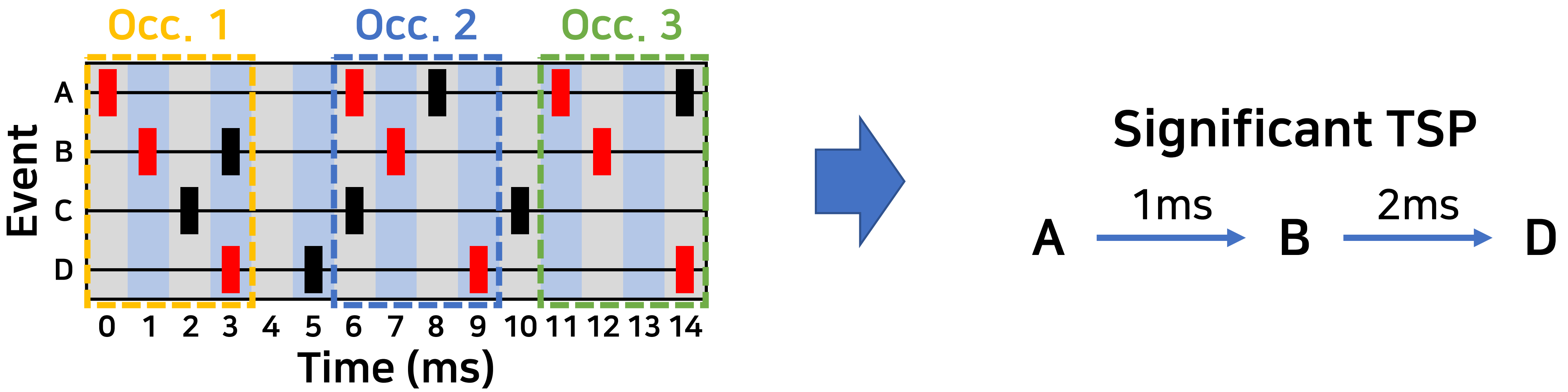
Group interactions are not observable
but signaled by temporal correlation

Problem Formulation

- **Given**: A temporal event sequence
- **Find**: Group interactions

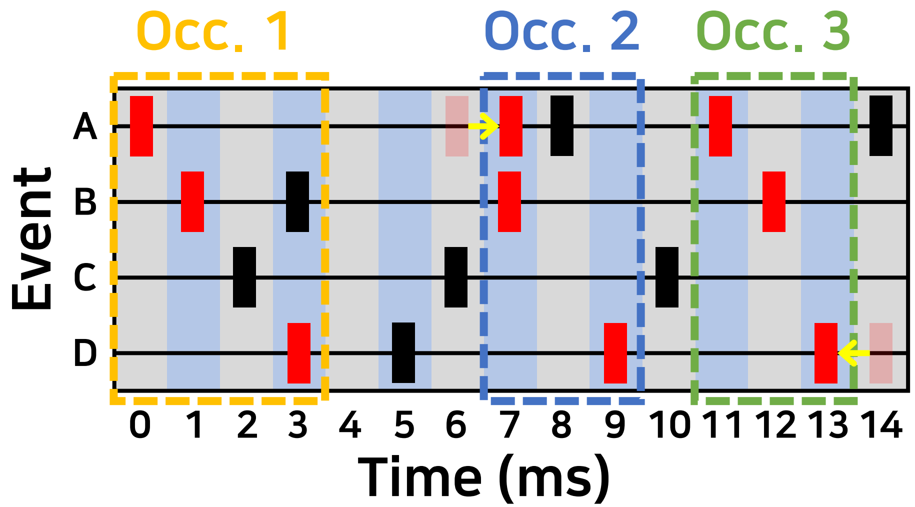
Group interactions are not observable
but signaled by temporal correlation

- **Given**: A temporal event sequence
- **Find**: Significant temporal sequential patterns



Challenge: Temporal Variations

- **Temporal Variations**: Time gaps in occurrences of a TSP might not be consistent across various instances in real-world data
  - E.g., Measurement errors or inherent system variability



TSP : $A \xrightarrow{1ms} B \xrightarrow{2ms} D$

Occ. 1: $A \xrightarrow{1ms} B \xrightarrow{2ms} D$

Occ. 2: $A \xrightarrow{0ms} B \xrightarrow{2ms} D$

Occ. 3: $A \xrightarrow{1ms} B \xrightarrow{1ms} D$

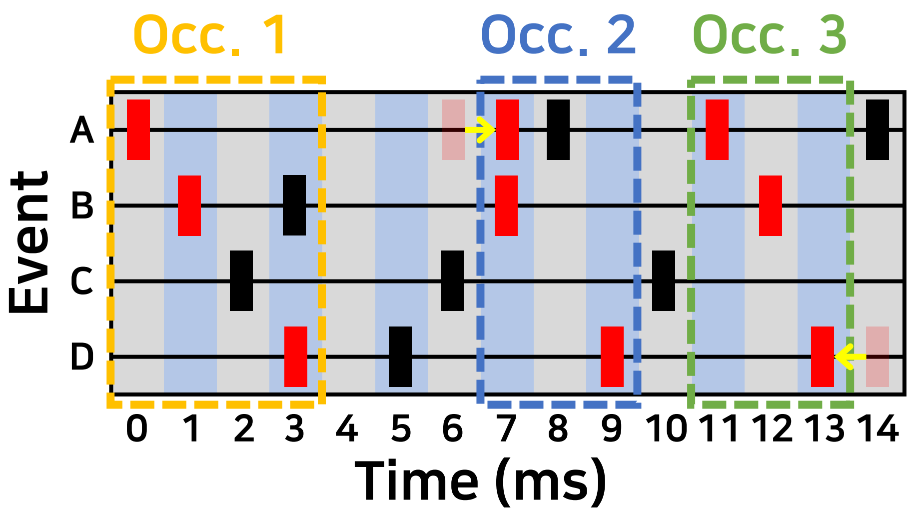
- Temporal variations present challenges for accurately discovering TSPs
  - However, previous methods do not explicitly consider temporal variations
  - **Our approach**: Introduce relaxed TSP concepts to handle temporal variations

Final Problem Formulation

- **Given**: A temporal event sequence with temporal variations
- **Find**: Group interactions on the ground-truth temporal event sequence

Group interactions are not observable
but signaled by temporal correlation

- **Given**: A temporal event sequence with temporal variations
- **Find**: Significant temporal sequential patterns on the ground-truth temporal event sequence



Significant TSP

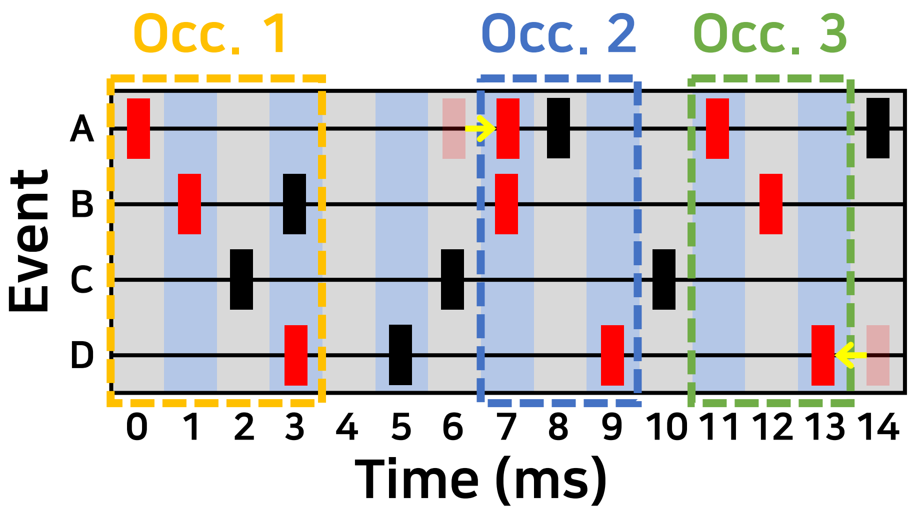
$A \xrightarrow{1ms} B \xrightarrow{2ms} D$

Overview of RASP

- **RASP** – **R**obust and resource-**A**daptive mining of temporal **S**equential **P**atterns
  - **Given** a temporal event sequence,
  - **Aims** to return the most significant TSPs
  - Key Idea
    - Novel concept: Relaxed TSP
    - Efficient search algorithm

Concepts: Relaxed TSP

- To improve robustness against temporal variations in the time gaps in a TSP, a relaxed TSP allows for a predefined level of time gap deviations
- **Relaxed TSP**: A sequence of events and relaxed time gaps
  - **Relaxed time gaps**: Intervals with a length of $2 \times \Delta$



An example of relaxed TSP (Δ=2ms)

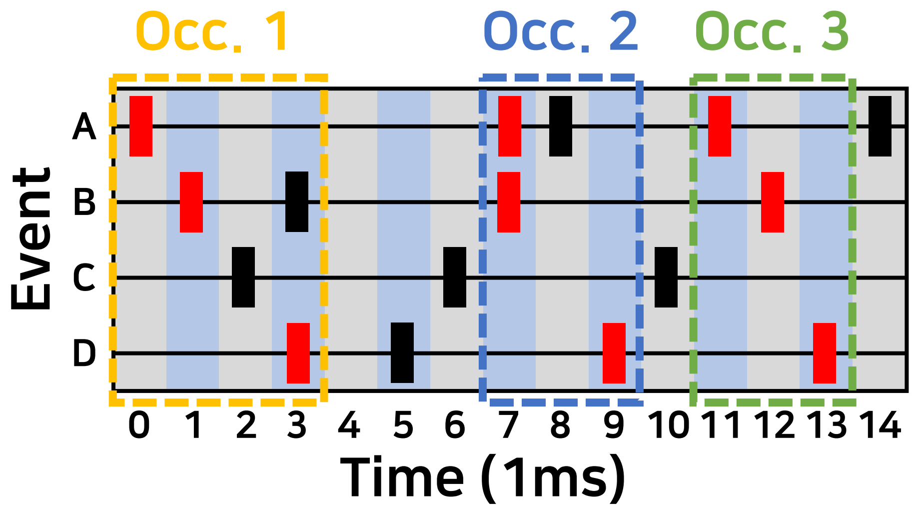
$$A \xrightarrow{2ms \pm \Delta} B \xrightarrow{2ms \pm \Delta} D$$

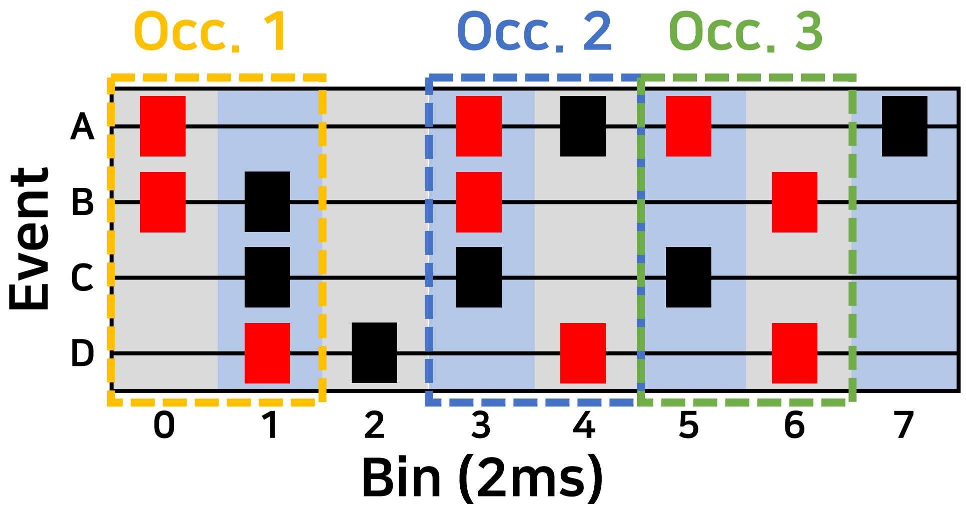
Occ. 1, 2, and 3 are all instances of the same relaxed TSP

Potential Alternative: Data Binning

- **Data Binning:** Events are grouped into small intervals, called bins

An example of binning (bin size=2ms)





Occ. 1, 2, and 3 are all instances of

$A \xrightarrow{2ms\pm\Delta} B \xrightarrow{2ms\pm\Delta} D$

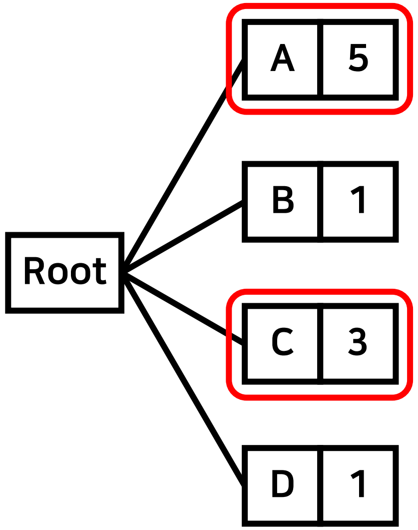
Occ. 1 and 2 are instances of $A \xrightarrow{0\text{ bin}} B \xrightarrow{1\text{ bin}} D$

while Occ. 3 is an instance of $A \xrightarrow{1\text{ bin}} B \xrightarrow{0\text{ bin}} D$

Details Search Algorithm

- **Challenge**: The number of TSPs is exponentially growing
- **Key Idea**: Beam search with resource-adaptive beam width
  - To retain a sufficient number of TSPs but without exceeding storage capacity, RASP automatically adjusts the threshold based on the resource-adaptive beam width

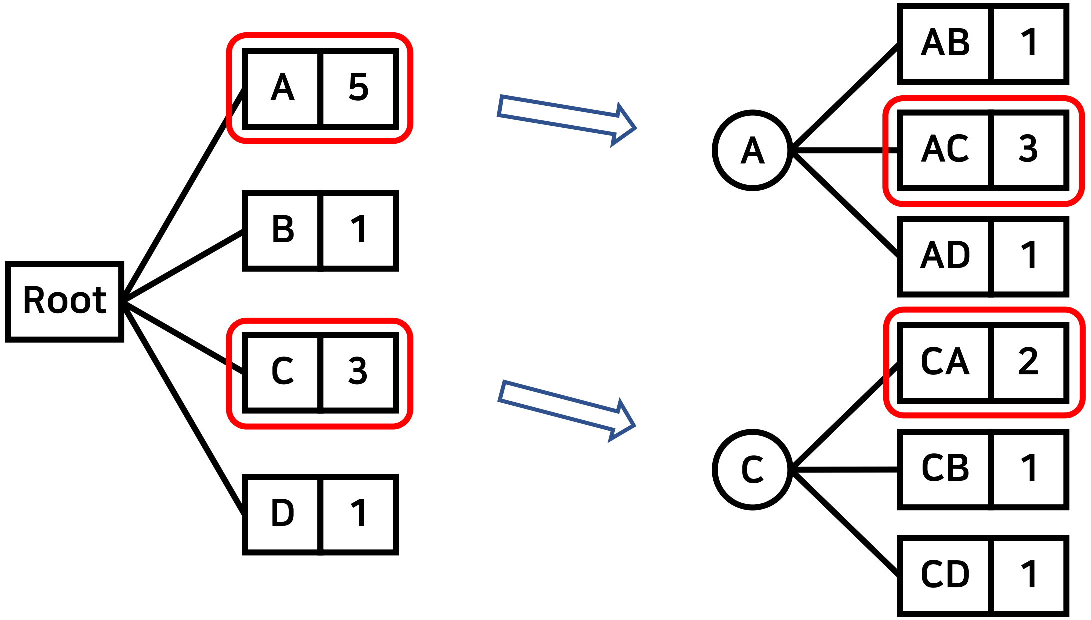
An example of beam search with resource-adaptive beam width
(Total resource: 4 × TSPs, Beam width: resource / 2 = 2)



Details Search Algorithm

- **Challenge**: The number of TSPs is exponentially growing
- **Key Idea**: Beam search with resource-adaptive beam width
  - To retain a sufficient number of TSPs but without exceeding storage capacity, RASP automatically adjusts the threshold based on the resource-adaptive beam width

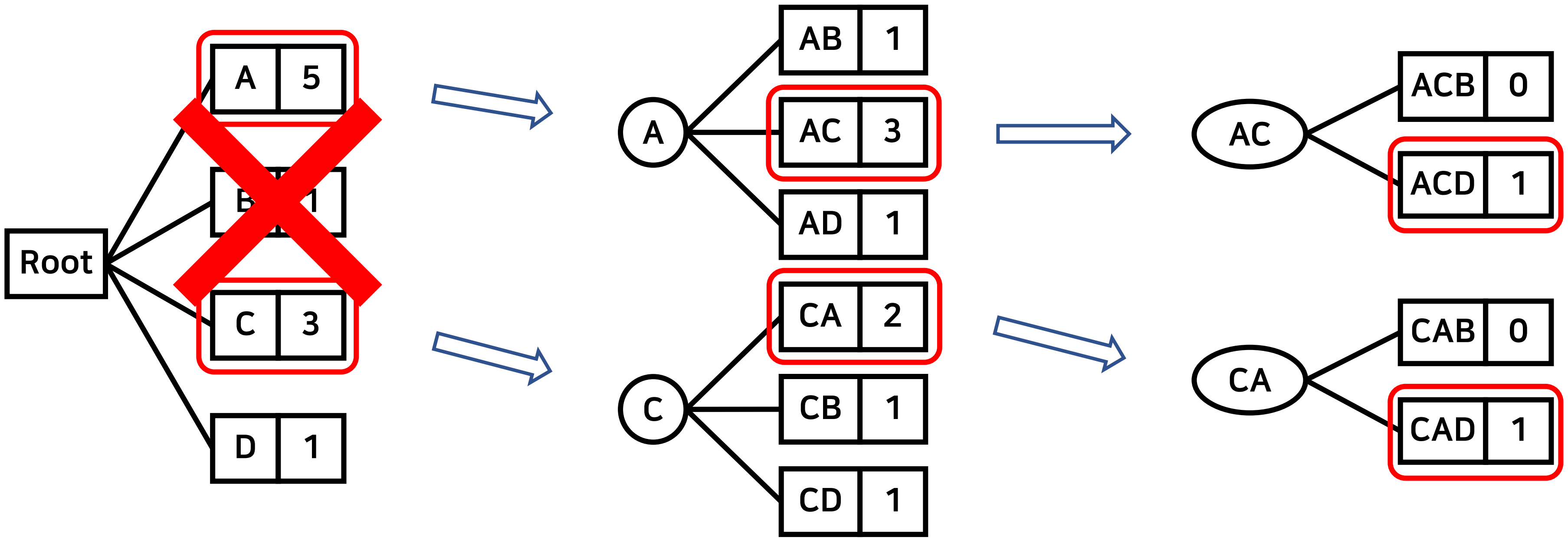
An example of beam search with resource-adaptive beam width
(Total resource: 4 × TSPs, Beam width: resource / 2 = 2)



Details Search Algorithm

- **Challenge**: The number of TSPs is exponentially growing
- **Key Idea**: Beam search with resource-adaptive beam width
  - To retain a sufficient number of TSPs but without exceeding storage capacity, RASP automatically adjusts the threshold based on the resource-adaptive beam width

An example of beam search with resource-adaptive beam width
(Total resource: 4 × TSPs, Beam width: resource / 2 = 2)

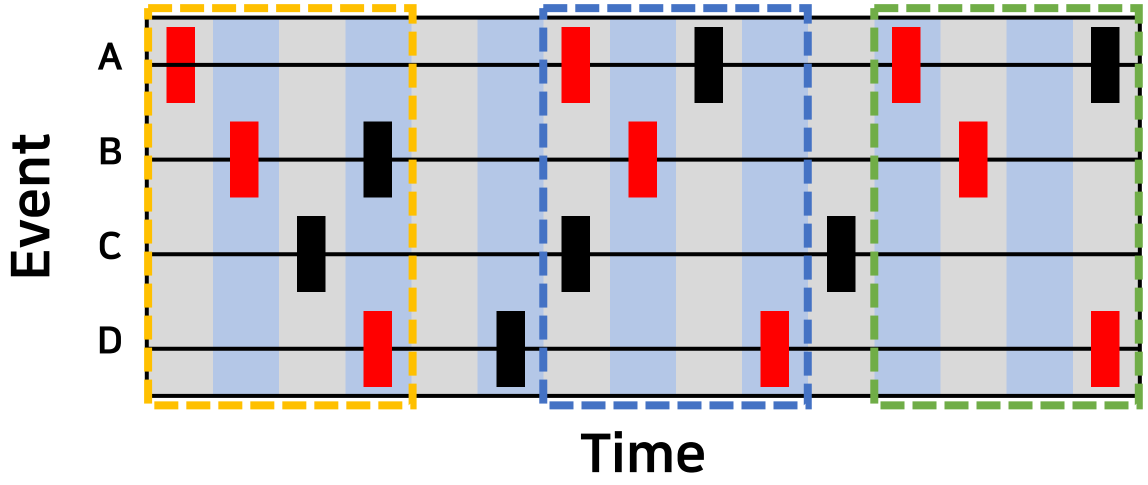


Potential Alternative: Predefined Fixed Threshold

- Predefined fixed thresholds can be used to determine the number of TSPs
  - E.g., significance level: p-value < 0.05
  - The number of TSPs is often sensitive to these threshold values, resulting in either an excessive or insufficient number of TSPs
  - Finding suitable threshold values, which vary across datasets, requires extensive trial and error

Synthetic Datasets

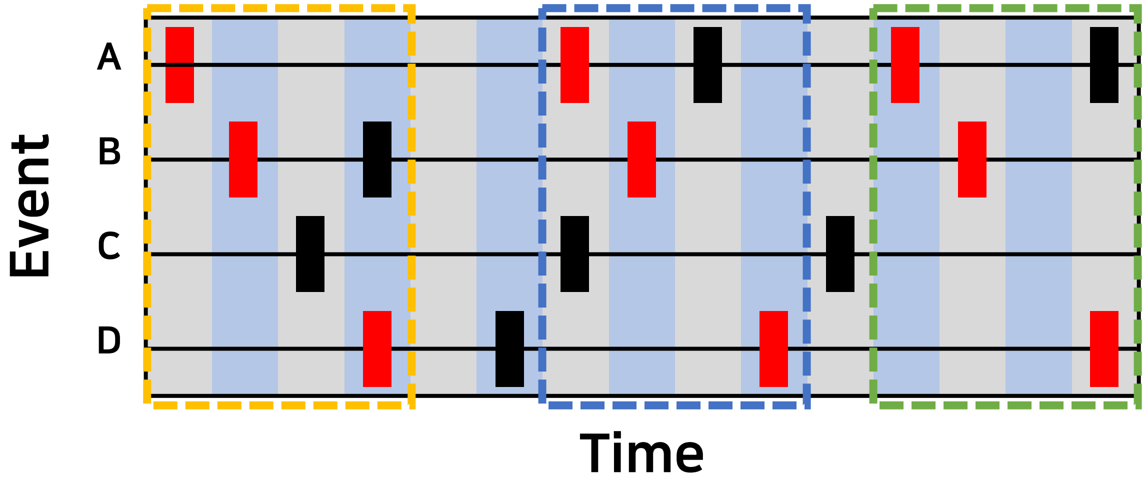
- Neuron activity datasets generated by the CN2 simulator[1]
  - Replicating real-world behaviors (e.g., temporal variations and probabilistic participation)
  - **Each event**: Spike of a specific neuron
  - **Ground-truth group interaction**: Functional groups of neurons



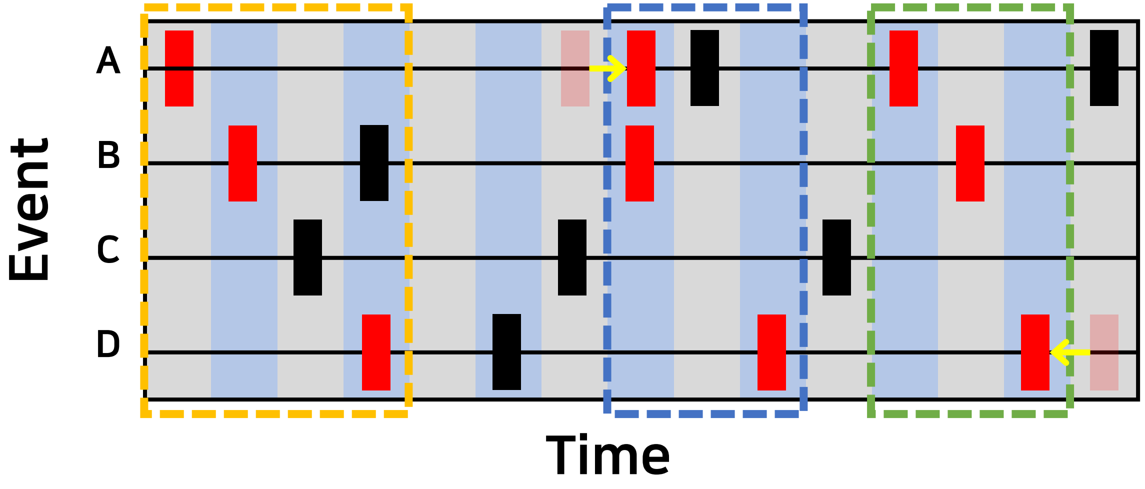
1. https://github.com/NICALab/CN2-Simulator

Temporal Variations

- **Variation-Free**: Without temporal variations
- **Variations**: With temporal variations
  - Zero-mean Gaussian noise with a predefined standard deviation to each time gap in the ground-truth TSP instances



Variation-Free



Variations

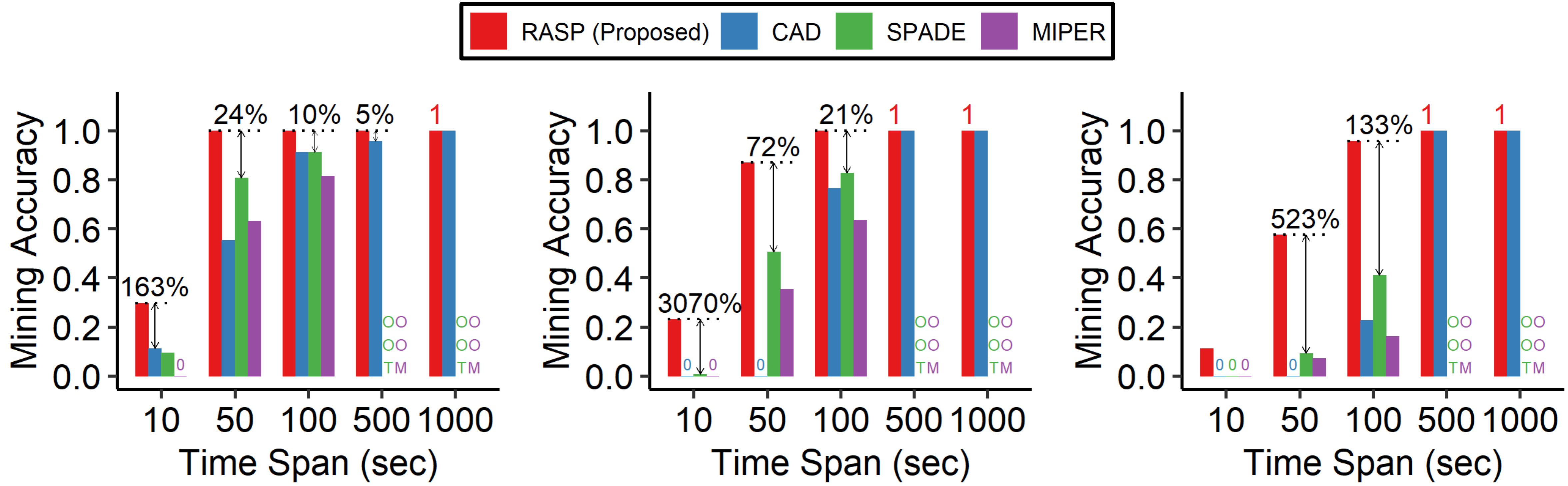
Experimental Settings

- Evaluation metric for accuracy: Normalized discounted cumulative gain (NDCG)
  - **NDCG@n** measures the quality of the top-n ranking by comparison with ground-truths
  - Ranges from 0 to 1, with higher values indicating better ranking quality
  - Note: The recall metric for accuracy produces similar results

- Competing methods
  - **CAD** [Russo et al., 2017]
  - **SPADE** [Torre et al., 2013, Quaglio et al., 2017]
  - **MIPER** [Ao et al., 2018]
  - These methods simply rely on data binning to handle temporal variations

Q1. RASP is accurate

- RASP outperformed all its competitors in accuracy across all settings



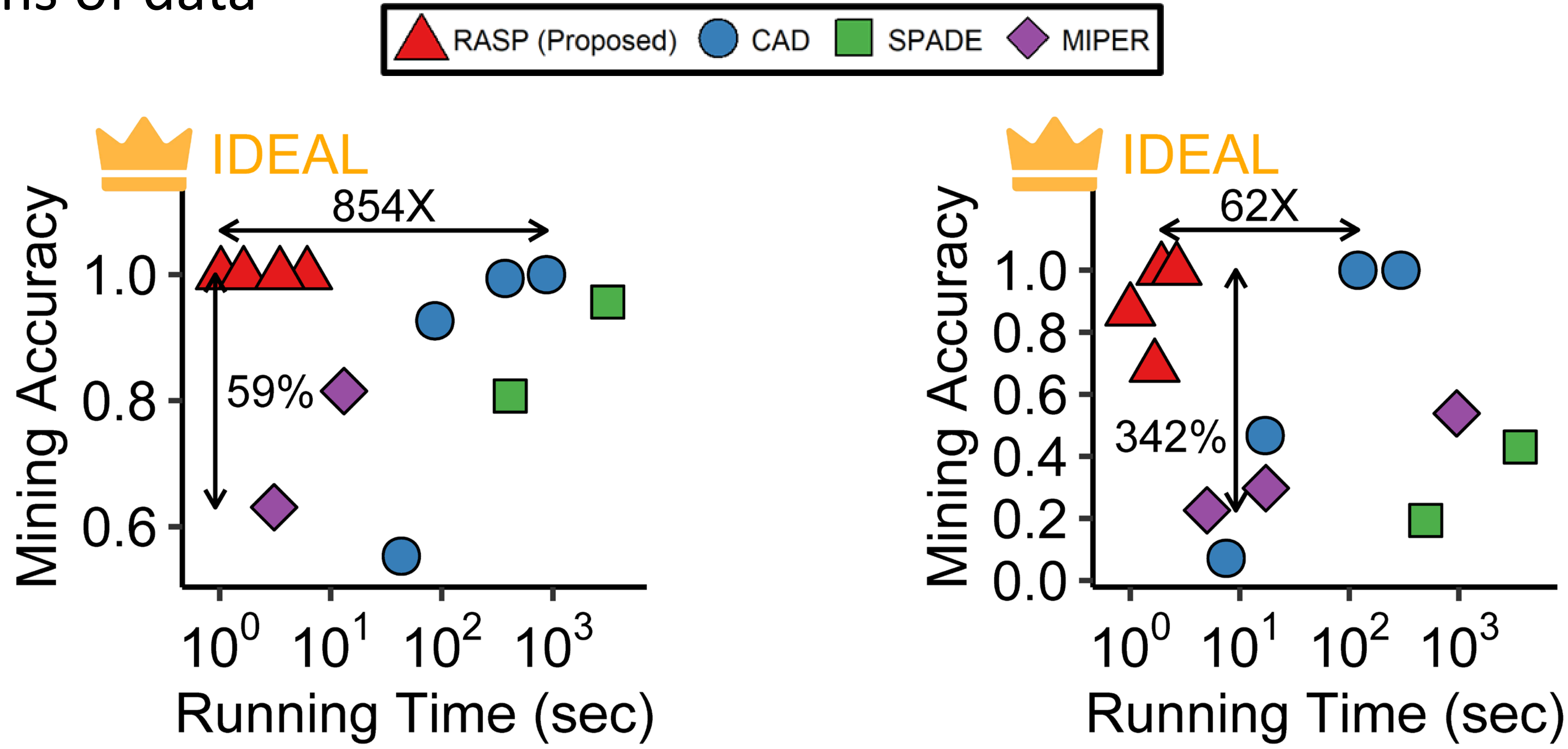
Variation-Free

Variations
(stdev: 10 ms)

Variations
(stdev: 20 ms)

Q2. RASP Gives a Better Speed-Accuracy Trade-off

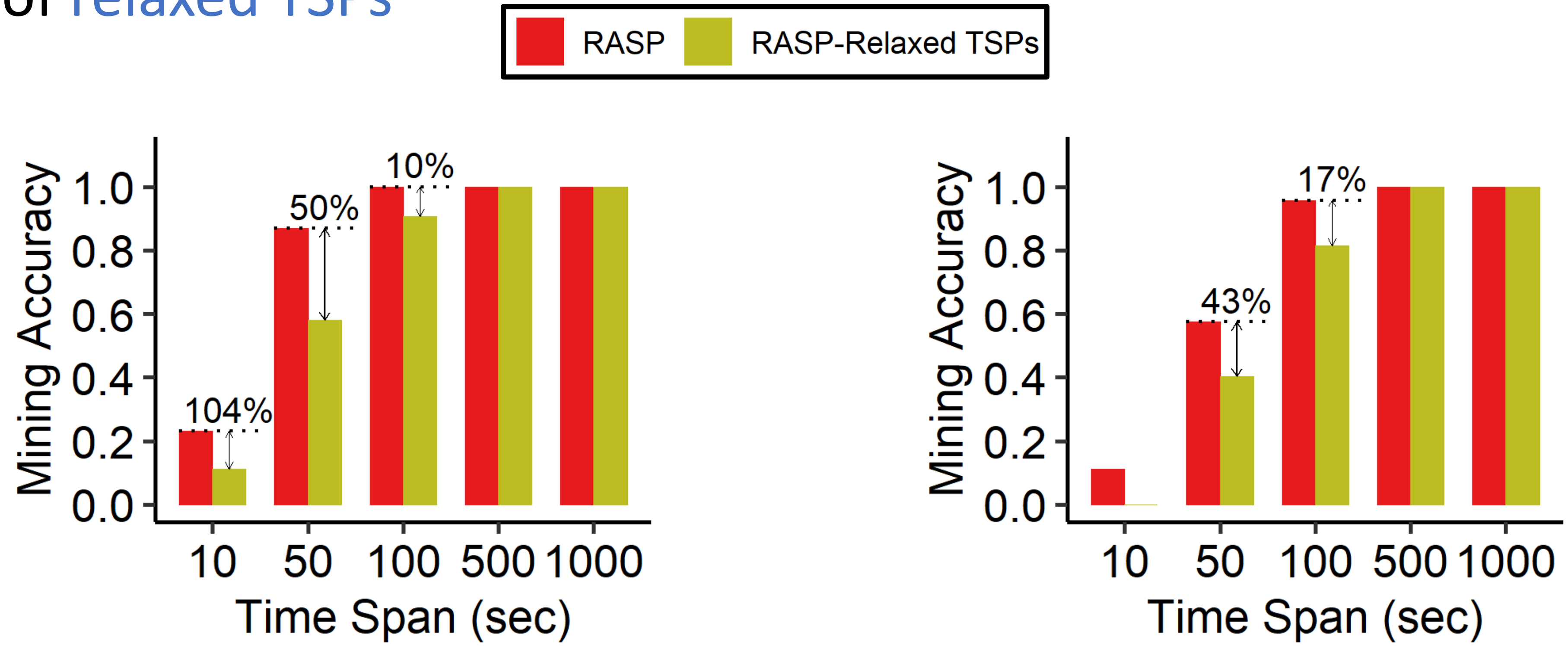
- RASP provided clearly better trade-offs than the other methods across all time spans of data



Variation-Free **Variations**

Q3. Relaxed TSP Contributes to Accuracy

- RASP tended to perform better compared to a variant without using the concept of relaxed TSPs

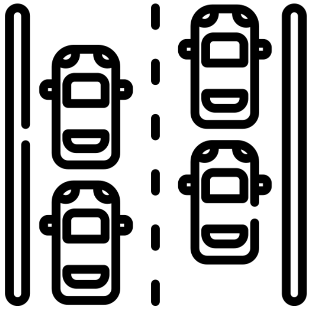


Variations (10 ms) **Variations (20 ms)**

Q4. Evaluation on Real-World Dataset

- Indirect evaluation on real-world datasets without ground-truth TSPs
  - Precipitation
    - **Each event**: Precipitation of $\geq 1\ mm$ accumulated over 15 min at each weather station
    - **Indirect accuracy**: % of reasonable TSPs (avg. distance $\leq 5\ km$)
  - Traffic congestion
    - **Each event**: Traffic volume of ≥ 500 vehicles accumulated over 15 min recorded at each sensor
    - **Indirect accuracy**: % of reasonable TSPs where (sum. distance $\leq 4\ km$)
  - Stock price fluctuation
    - **Each event**: Daily return rate of $\geq 5\%$ or $\leq -5\%$

Precipitation

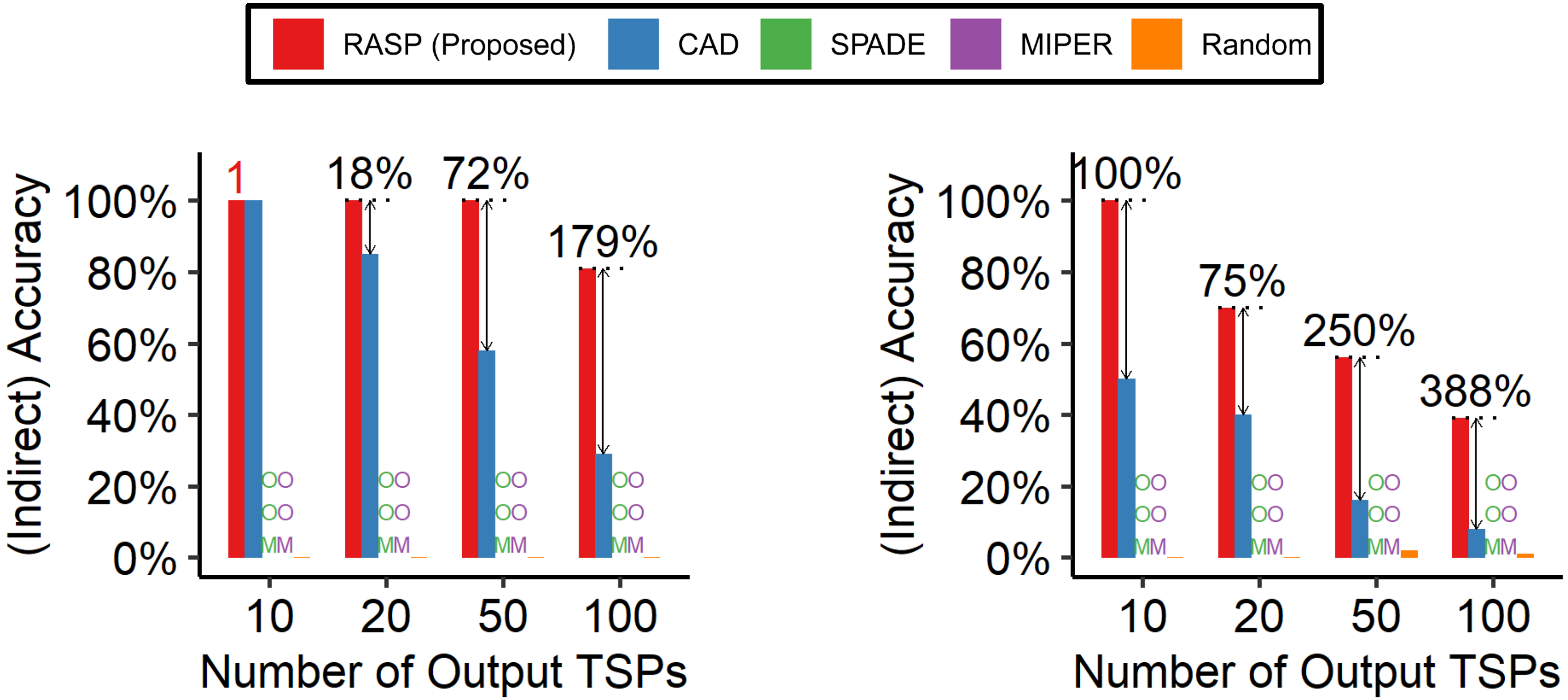


Traffic

Stock

Q4. Evaluation on Real-World Dataset

- RASP consistently achieves the highest (indirect) accuracy



Precipitation **Traffic**

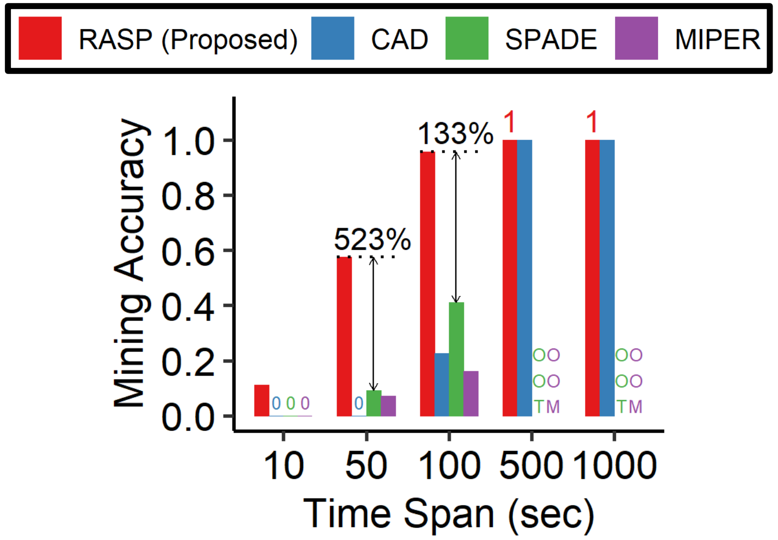
Q4. Evaluation on Real-World Dataset

- RASP identified frequent TSPs with stock price fluctuation events, revealing:
  1. **Sector-based** patterns: Stocks from the same sector
     - Most frequent TSP: Engineering & Construction sector
     - 2nd most frequent TSP: Shipbuilding & Offshore Engineering sector
  2. **Affiliate-based** patterns: Stocks from the same corporate affiliates
     - 3rd most frequent TSP: Doosan group

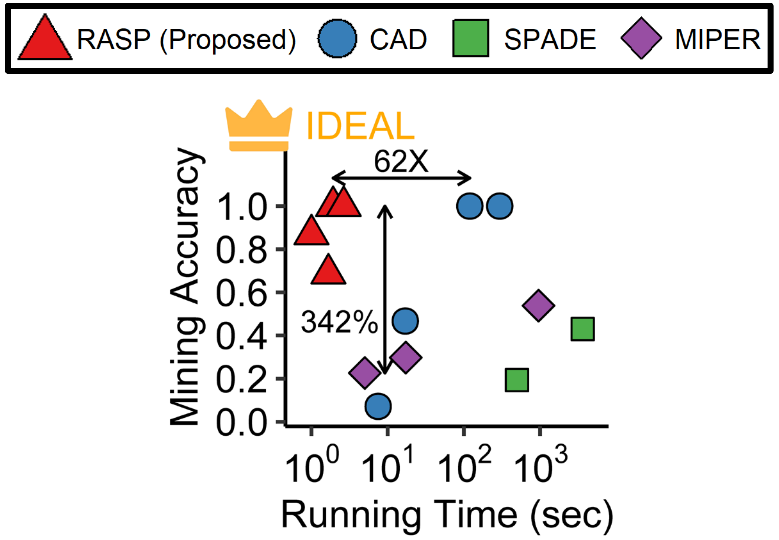
Conclusion

- We proposed **RASP**, an algorithm for mining significant TSPs in a sequence of temporal events with temporal variations, which incorporates:
  - A novel concept of relaxed TSPs for handling temporal variations
  - Resource-adaptive automatic hyperparameter tuning for enhancing usability

Accurate and Robust



Speed-Accuracy Trade-Off

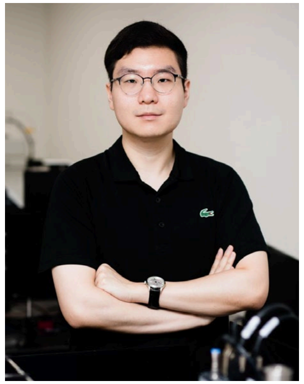


Source code and datasets are available at https://github.com/jin-choo/RASP

EDBT ICDT BARCELONA 2025

KAIST

RASP: Robust Mining of Frequent Temporal Sequential Patterns under Temporal Variations


<u>Hyunjin Choo</u> Minho Eom Gyuri Kim Young-Gyu Yoon Kijung Shin