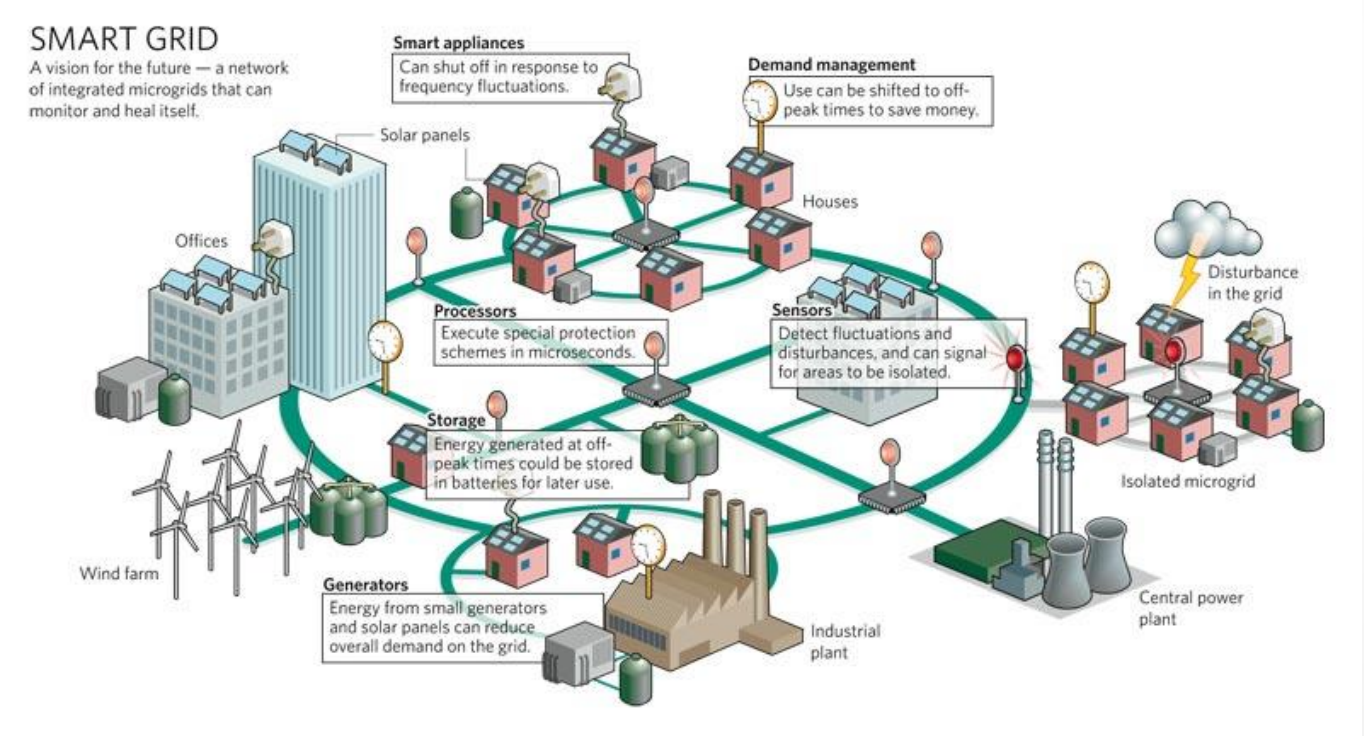
# Online Temporal-Spatial Analysis for Detection of Critical Events in Cyber-Physical Systems

Zhang Fu, Magnus Almgren,
Olaf Landsiedel, Marina Papatriantafilou
*Chalmers University of Technology*

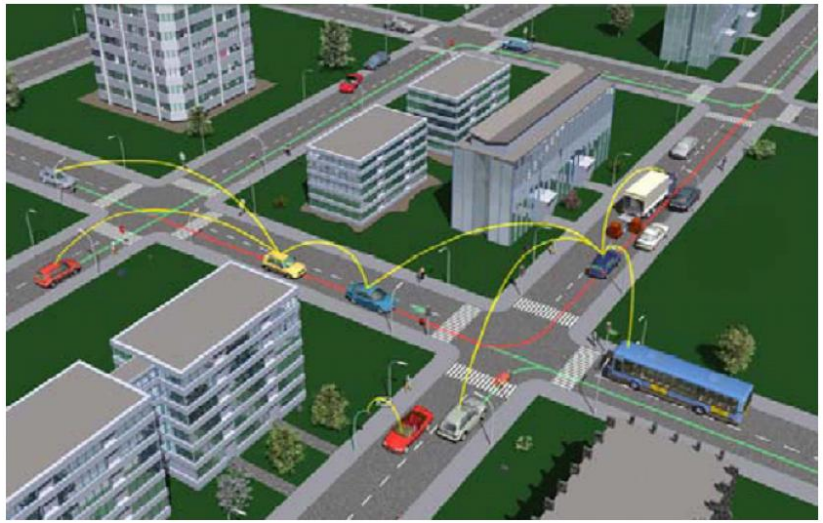
# Example Cyber-Physical Systems

- **Smart grid**:
  monitoring electricity distribution networks, reporting faults and critical events, controlling smart devices, etc.



# Example Cyber-Physical Systems

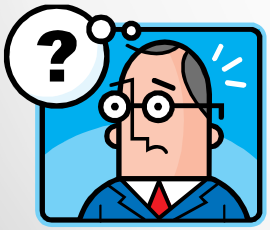
- **Traffic monitoring system**:
  communicating with vehicles, monitoring traffic conditions, reporting jams and accidents and scheduling traffic lights, etc.



http://www3.imperial.ac.uk/pervasivesensing/projects/trafficmonitoring

# Problem

- **Hundred of thousands of sensors** deployed
  - Smart meters in a city
    (often called the **A**dvanced **M**etering **I**nfrastructure, AMI)
  - Cars on the road
- **Considerable number of messages** sent from each sensor
  - Transit faults
  - Instability
  - Noise
  - Real Data
- **Goal:**
  - Identify critical events
  - Efficiency

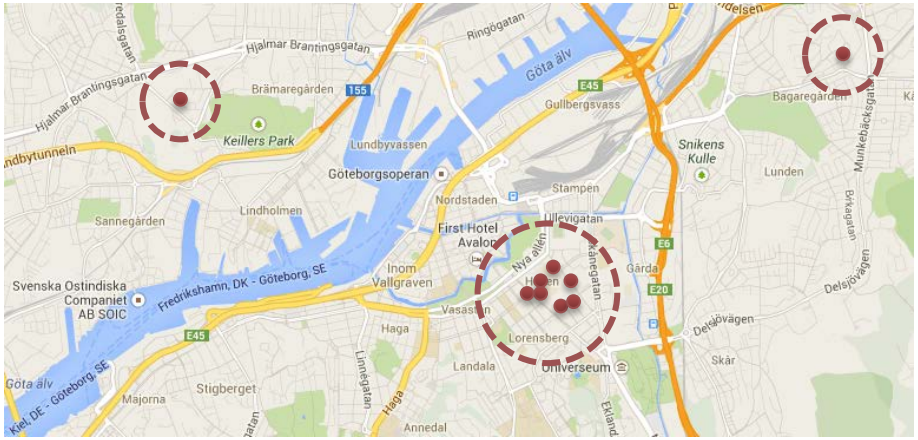



*Messages from different sensors may indicate the same event, how can we find it?*

# Example of critical event (from AMI)

**Significant power outages happen in the Grid**

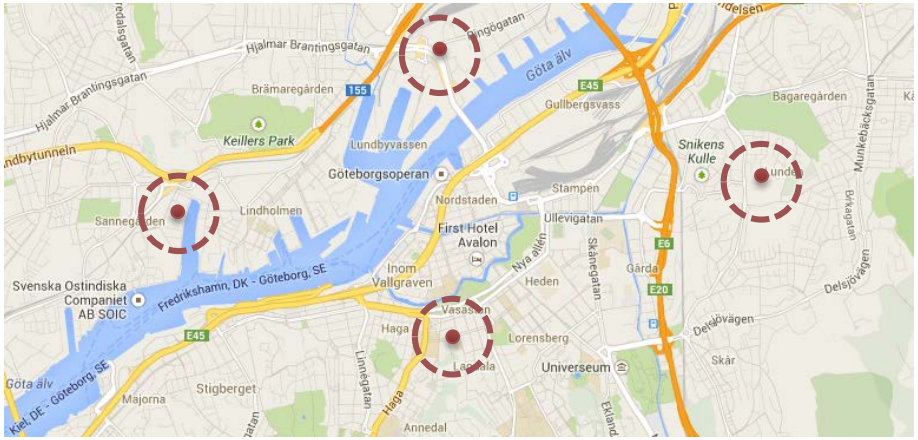
- Significant number of power failure messages
- Short time interval




# No critical event (from AMI)

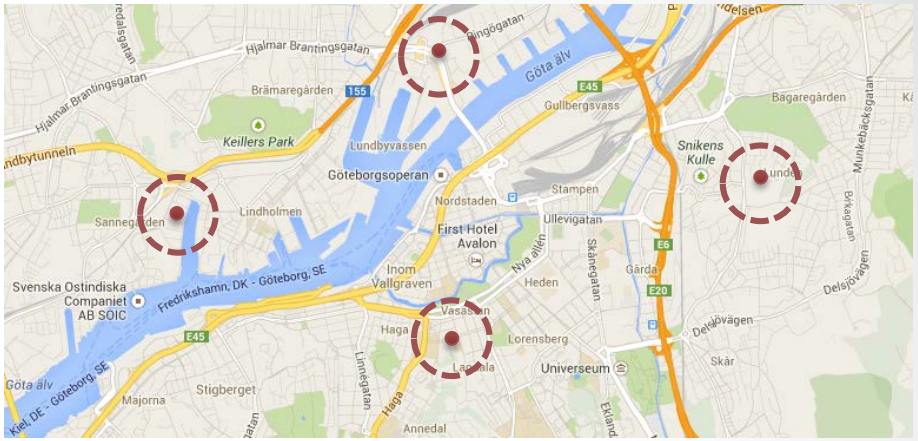
**The electricity network is functioning normally**

- Random power failure messages
- Short time interval

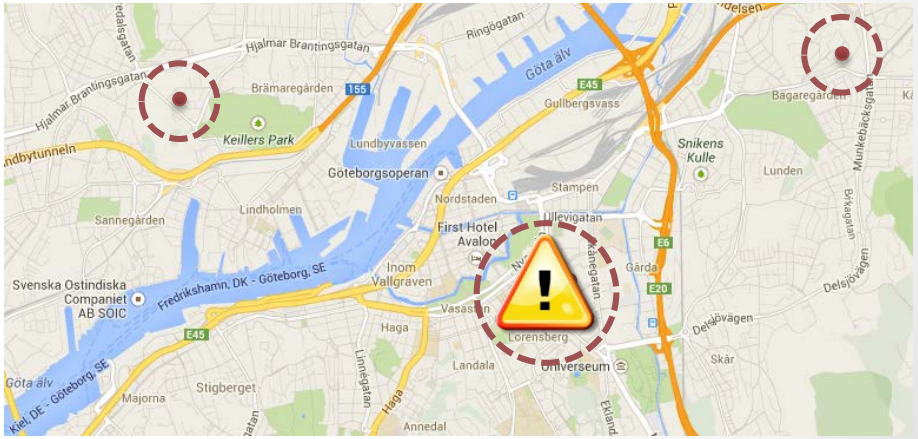


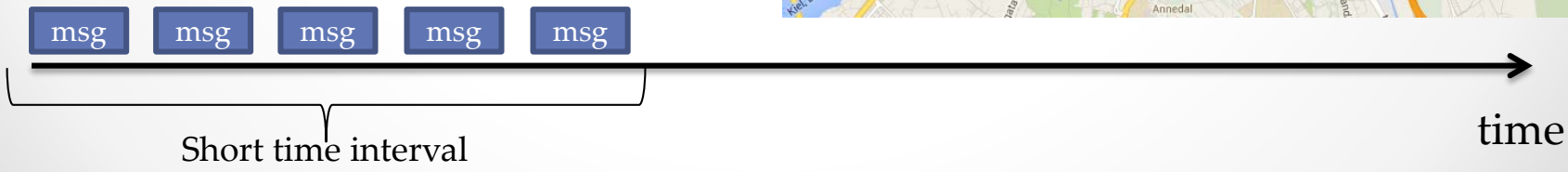

# Comparison

- **Similar properties**
  - Lot of power failure messages
  - Short time interval




- **Differences**
  - Locality

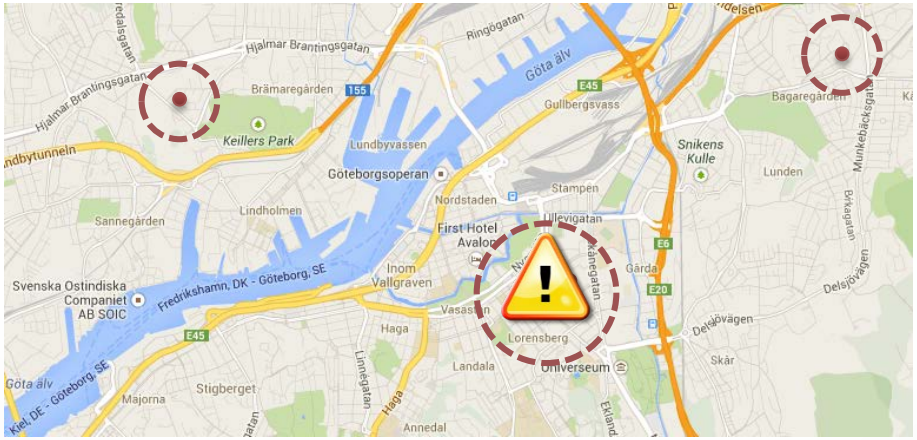


# Motivation (from AMI)

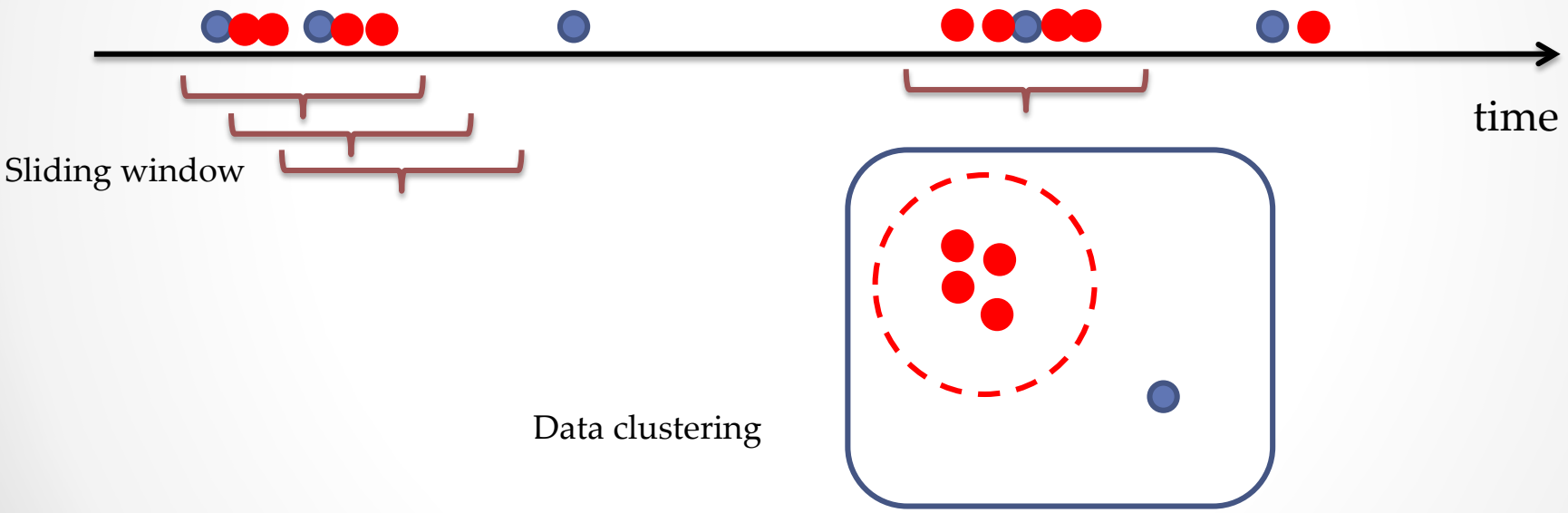
- **Temporal properties**


- **Spatial properties**



# Clustering Data Stream over Sliding Windows

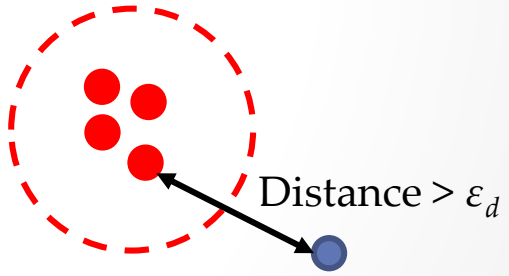
- Sliding window helps to filter out time-unrelated records
- Clustering helps to find spatial- related records



# Single-Linkage Clustering (SLC)

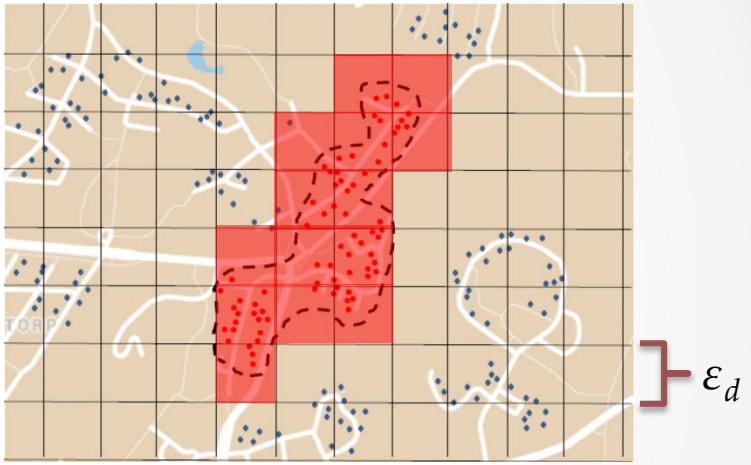
- Model
  *Single-linkage clustering with stop-condition $\varepsilon_d$.*
  - Detecting clusters with arbitrary shape.
  - Does not need to define densities as in th… density-based clustering algorithm… (e.g. DBSCAN).
- Algorithm
  - Assign… …clusters)
  - *… distance < $\varepsilon_d$, merge th… and delete the old two*
  - *Re… previous until all distance > $\varepsilon_d$ or there is only one cluster*
- Both space and time complexity of doing single-linkage clustering is $O(n^2)$.

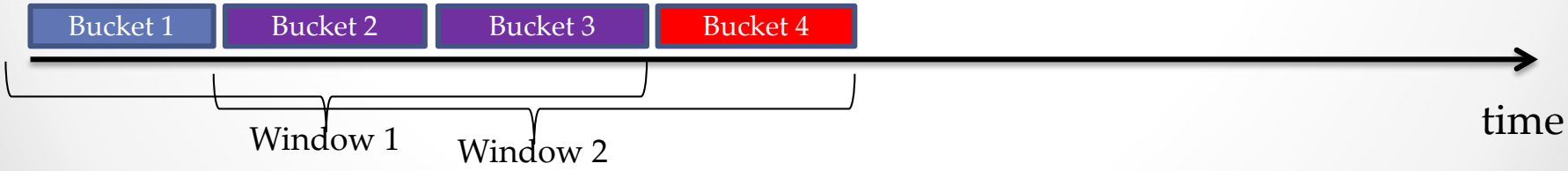
Not acceptable for quick processing of massive number of records.



# Grid-based Single-Linkage Clustering (G-SLC)

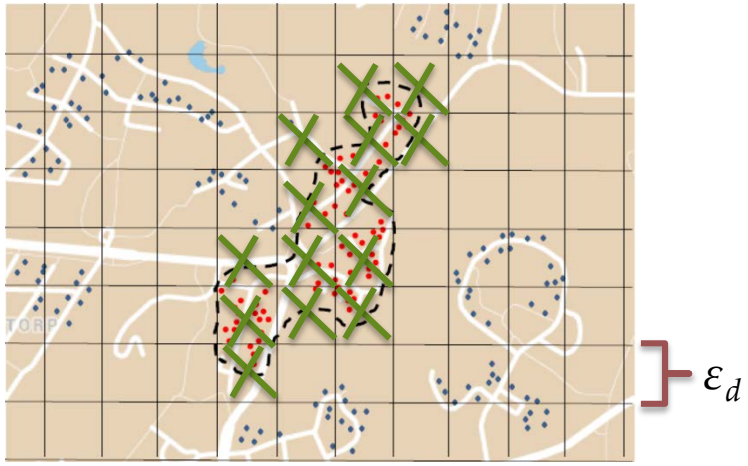
- Aggregate records by cells in the grids
  - The weight of a cell is the number of records in the cell in the current window.
  - Only the cells containing records are maintained in memory. These cells are called ***active cells***.
  - Adaptive to the sliding window, the weight of a cell is kept in buckets.



# Grid-based Single-Linkage Clustering (G-SLC)

- Clustering active cells
  - The cell size is $\varepsilon_d$.
  - Go through all the active cells and join their active neighbor cells into the same clusters.

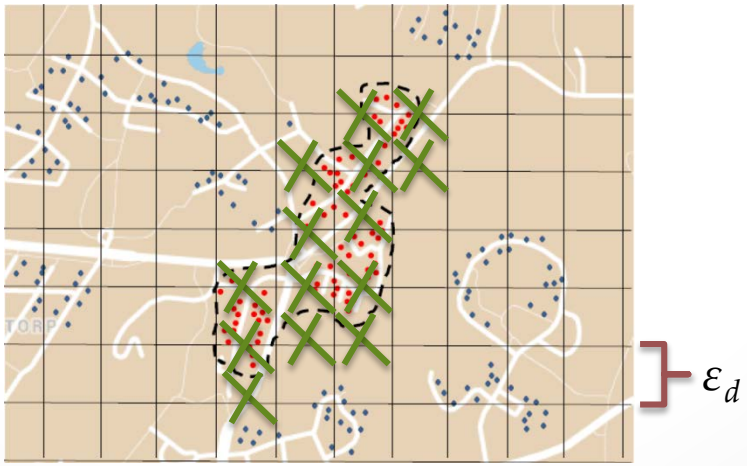


# Grid-based Single-Linkage Clustering (G-SLC)

- Significant clusters
  - Not a trivial job
  - Method should be efficient
  - Defining a significance function

$$\text{significance}(\text{Cluster}_i) = \frac{\#\ of\ records\ in\ Clusteri}{\#\ of\ records\ in\ the\ window}$$

  - Using threshold $\theta_s$



# Complexity and Clustering Performance

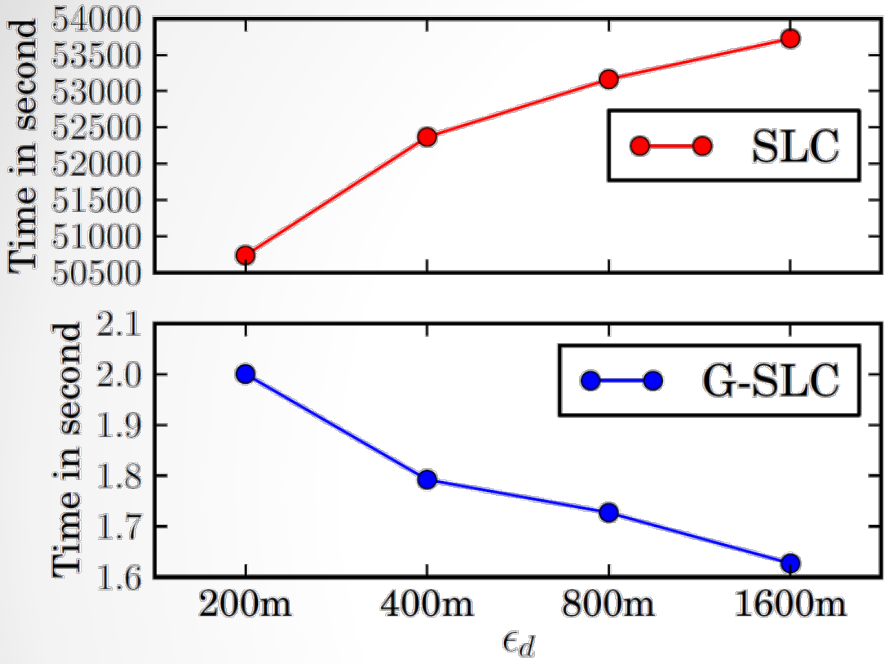
- Time complexity of mapping a record to a cell
  O(1)
- Time complexity of clustering active cells
  O(N), where N is the number of active cells
- Space complexity is O(N)
  - Each active cell only maintains a constant number of buckets
- If two records belong to the same cluster in SLC, then they belong to the same cluster in G-SLC
  - G-SLC does not miss clusters compared with SLC

# Performance Evaluation

- Advanced Metering Infrastructure (AMI) of an European city with roughly 600,000 inhabitants and around 300,000 smart meters.
  - Dataset contains all power failure messages from meters during 2012.
  - Each such message contains the timestamp and the location coordinates.

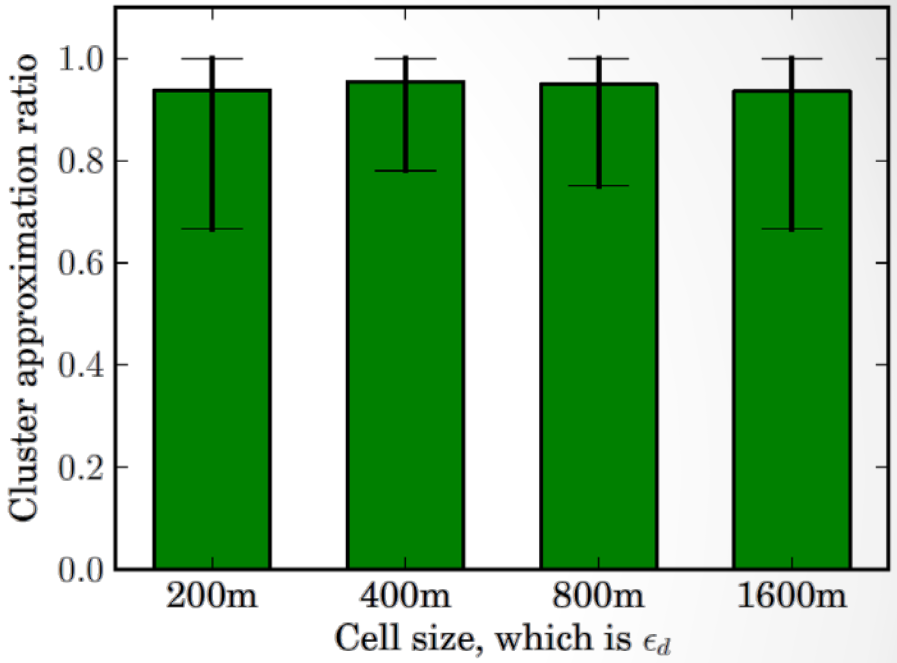
- Comparison with baseline algorithm SLC.
  - **Computation time**: a comparison between the performance of SLC and G-SLC to process the dataset.
  - **Cluster approximation**: a comparison of the number of clusters in SLC with G-SLC.

# Performance Evaluation




## Dramatic improvement in running time

G-SLC is approximately 2,500 times faster than SLC.




## Similar clusters

However, G-SLC still performs similarly to SLC in regards to the number of clusters created. Cluster approximation ratio:

$$\frac{\#\,\text{clusters G - SLC}}{\#\,\text{clusters SLC}}$$

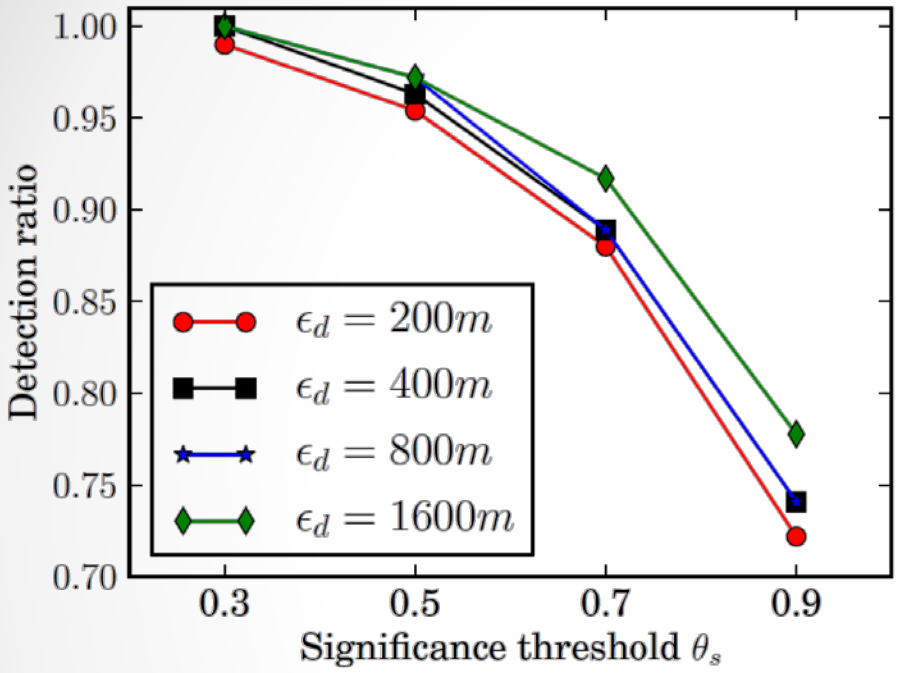
# Detection Evaluation

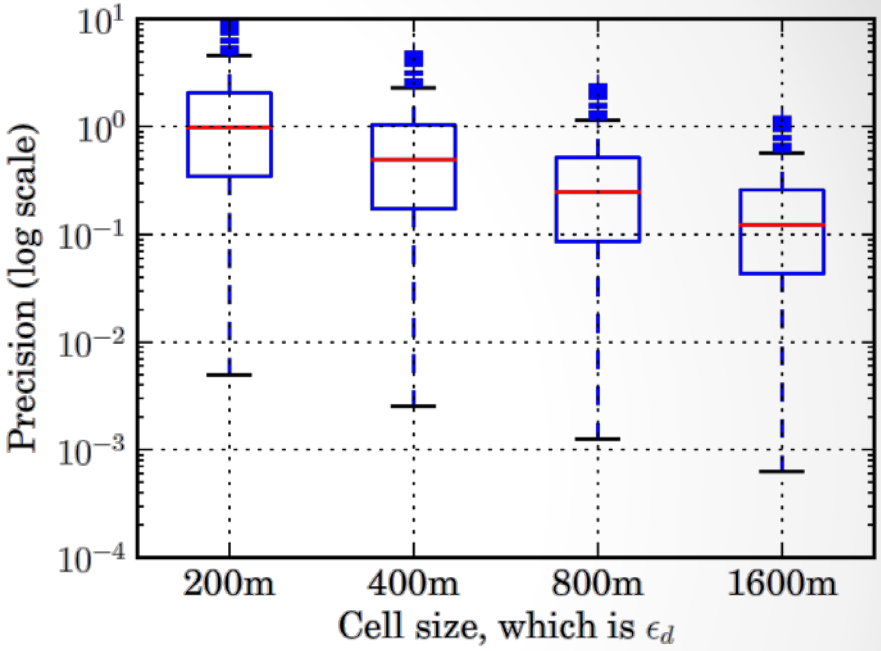
- Ground truth (reference events):
  registered power outages by the call center of the electricity company.

- Measures:
  - *Detection ratio: proportion of reference events also detected by G-SLC.*
  - **Affected by a threshold to filter noise and cell size.**
    **In the best case, we detect close to 100% of the events.**

  - *Detection precision: how well can we identify outages given different cell sizes?*
  - **Quite good but dependent on cell size and size of outage. For the real world data, a cell size of 400m give good accuracy.**

# Detection Evaluation




By choosing appropriate parameters, G-SLC can detect critical events with high detection ratio.

The significance threshold is helping to get the most significance cluster (see the paper for detail)

Enlarge the cell size will decrease the detection precision, the location of the outages given by G-SLC may be very coarse

# Conclusion

- Model the problem of detecting critical events in Cyber-Physical system as single-linkage clustering over sliding windows.
- Propose new algorithm
  - G-SLC for clustering data streams in a time-space efficient way.
- Evaluation of G-SLC with real-world data:
  - Compared to the baseline SLC, we show that it is significantly faster (up to 2,500 times) but with similar cluster behavior.
  - Compared to data collected by a call center for power outages, we show that we detect significant outages and can locate them geographically.
- Show that G-SLC can help to detect critical events with high detection ratio and precision.

# Future Work

- Similar clustering could be done over quality parameters of the grid
  - Voltage
  - Network stability
- Identification of power restore conditions