INSTRUCTOR: DR. KANGAVARI

PRESENTER: M. DOOSTI LAKHANI

Anomaly In Stream

ROBUST RANDOM CUT FOREST BASED ANOMALY DETECTION ON STREAMS

What We Will Cover

1. Intro
   1. Why stream data
   2. Main challenge
2. Study case: Anomaly detection
   1. What is anomaly?
   2. Importance of anomaly detection
   3. Synopsis of input data
   4. Updating synopsis efficiently
   5. Ensemble of RRCTs
   6. Viability

1 Intro

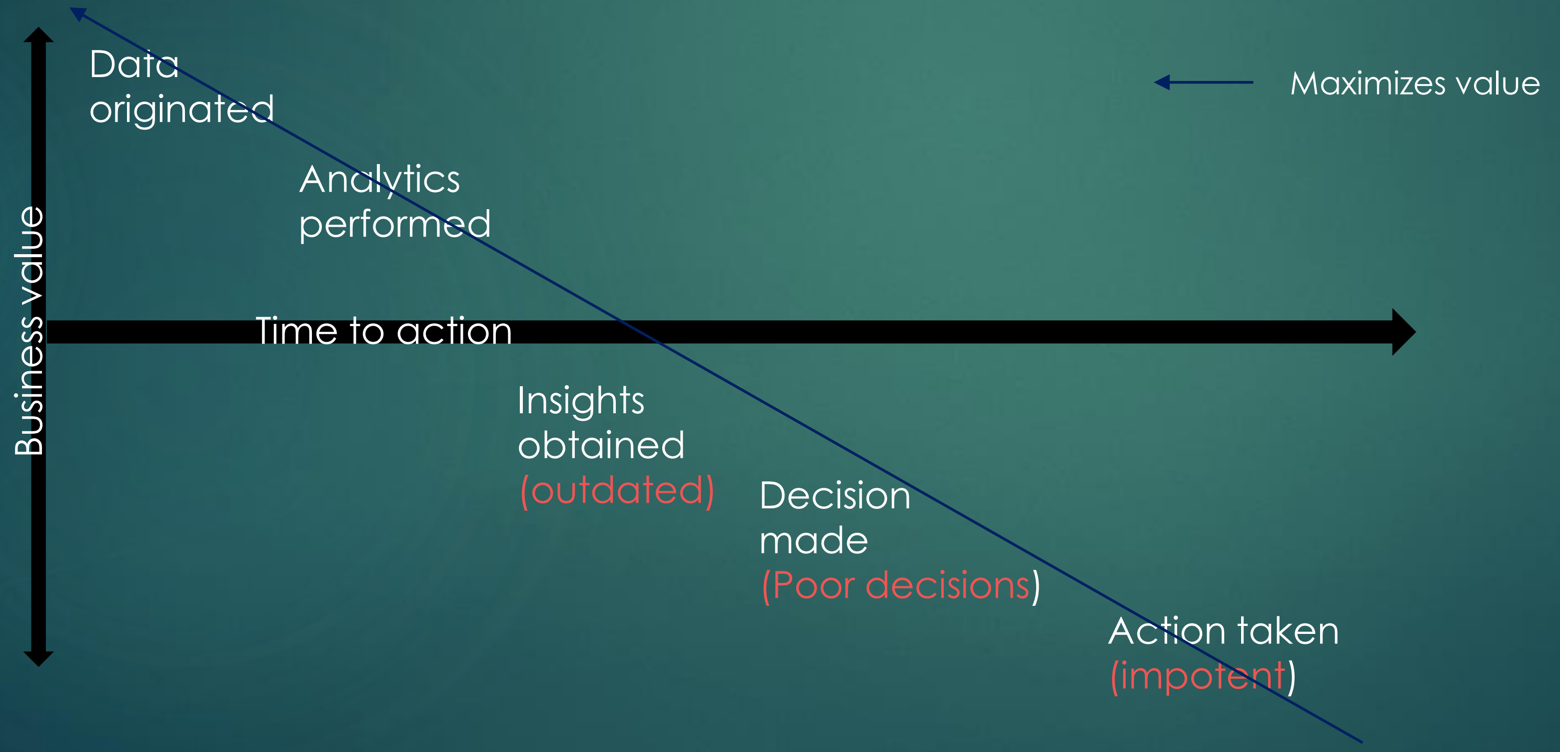
1. Why stream data
2. Main challenge

1.1 Why Stream Data

- All data originates in real time
  - Image segmentation, Language modeling
- Emerging explosion of IoT and internet, rejuvenated the well-studied problems such as anomaly detection
- We are dealing with it in everyday tasks

1.2 Main Challenge

- **Insights are perishable**
- Batch analytics operations take too long



1.2 Main Challenge cont.

- What we want:
  - Ingest data as it it's generated
  - Process data on the fly
  - Real time machine learning
- Robust Random Cut Forest

2 Study Case

1. What is anomaly?
2. Importance of anomaly detection
3. Synopsis of input data
    1. Algorithm
    2. Examples
    3. Definition of Anomaly in RCF
    4. Classic Shortcomings
4. Updating synopsis efficiently
5. Viability

2.1 What Is Anomaly?

- Anomaly is an observation that diverges from otherwise well-structured data
  - Outlier, exception or anything that deviates from normal pattern
- Model based perspective: A point is anomaly, if it increases the complexity of model
  - In case of trees, creating new leaves in early stages
  - Far from what has been learned, easier isolation

2.2 Importance of Anomaly Detection

- Anomalies need to be responded fast and accurately
  - A anomalous behavior in a patience gathered from their smart watch
  - Finding a failure in network systems
  - …

2.3 Synopsis of input data

1. Algorithm
2. Examples
3. Definition of Anomaly in RCF
4. Classic Shortcomings
5. Ensemble of RRCTs
6. Viability

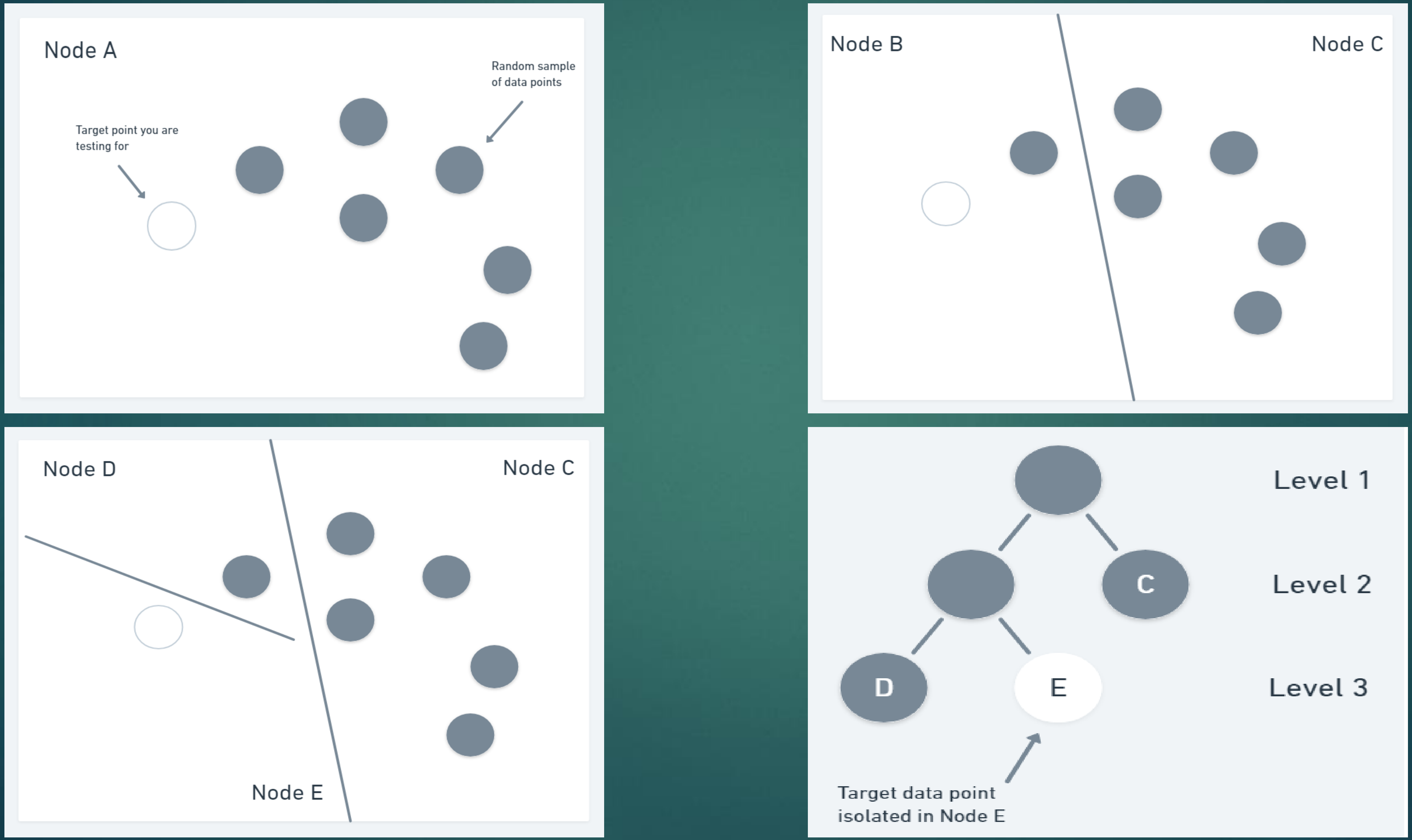
2.3.1 Algorithm

- RRCT
  - Unsupervised
  - Very fast on high dimensional data
- A **R**obust **R**andom **C**ut **T**ree on point set S
  - Choose a random dimension proportional to:

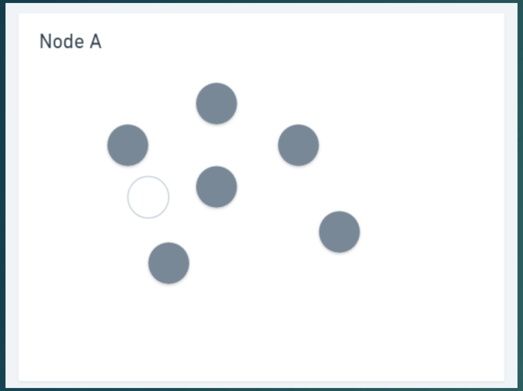
    $$\frac{l_i}{\sum_j l_j}$$

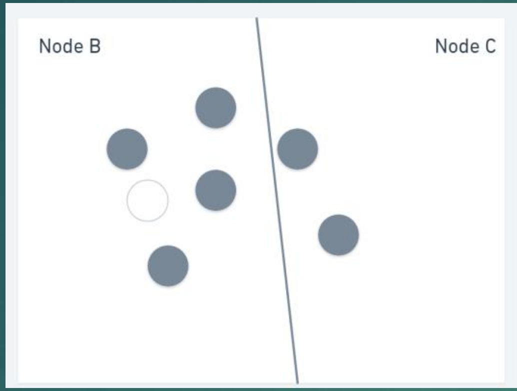
    where $l_i = \max(x_i) - \min(x_i)$
  - Choose $X_i \sim uniform(\min(x_i), \max(x_i))$
  - Let $S_1 = \{x | x \in S, x_i < X_i\}$ and $S_2 = \frac{S}{S_1}$ and recurse on S_1 and S_2

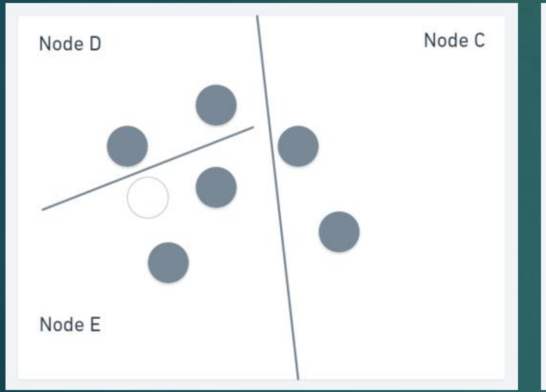
2.3.2 Examples

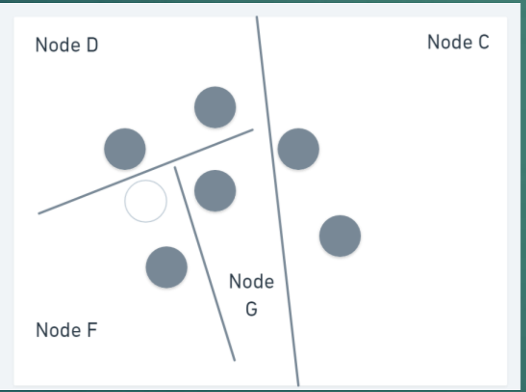


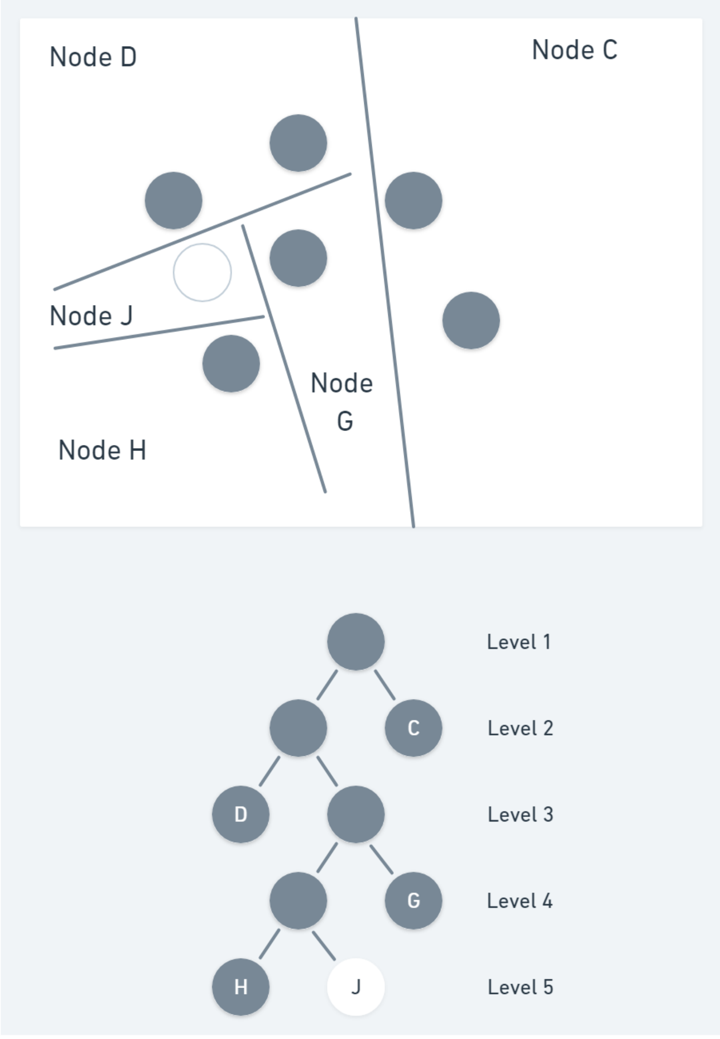
2.3.2 Examples cont.











2.3.3 Definition of Anomaly in RCF

- Anomaly points are isolated much faster and they will be on top of the tree
- The points near the root will get higher score
- The score is distance based, so the points far from normal clusters get higher score (nearer to root)
  - So a point will be anomaly if it increases the size of tree profoundly.
- Same idea has been used for test
  - If test node is near to root, then it is probably an anomaly
- If a point is far in from data in N-dim data, it will be as far relatively in RCF

2.3.4 Classic Shortcomings

- Classic approaches such as thresholding peaks for detecting anomalies were not successful
  - Could not be adopted from one task to another
  - Could not fit the nature of stream data
  - It needed expert which is against the automation!

2.4 Updating synopsis efficiently

- If we delete a node containing a isolated point x, and its parent, then the resulting tree has the same probability if it is being drawn without x
- By extending previous theorem, we can construct a tree without x but by adding it after construction.
- Typically, if we build a tree and insert a node and then delete it again, the result will be almost the same and preserves the distribution.
- This enables the adaption in stream learning

2.4 Updating synopsis efficiently cont.

- We can maintain a random tree over s sample set even as the sample is updated dynamically for streaming data using sublinear update time $O(d|S|)$
- For sliding window over data, we can think of removing a node and adding other nodes.
  - Removing nodes can be done by deleting nodes with lower priority
  - As we do this uniformly, we can think of last recently used as the choice.
- Based on the theorem in previous slide, we can efficiently answer this question that what would happen if we add arbitrary point p but constructing its tree.
  - Inserting is like constructing the tree with p from the beginning.

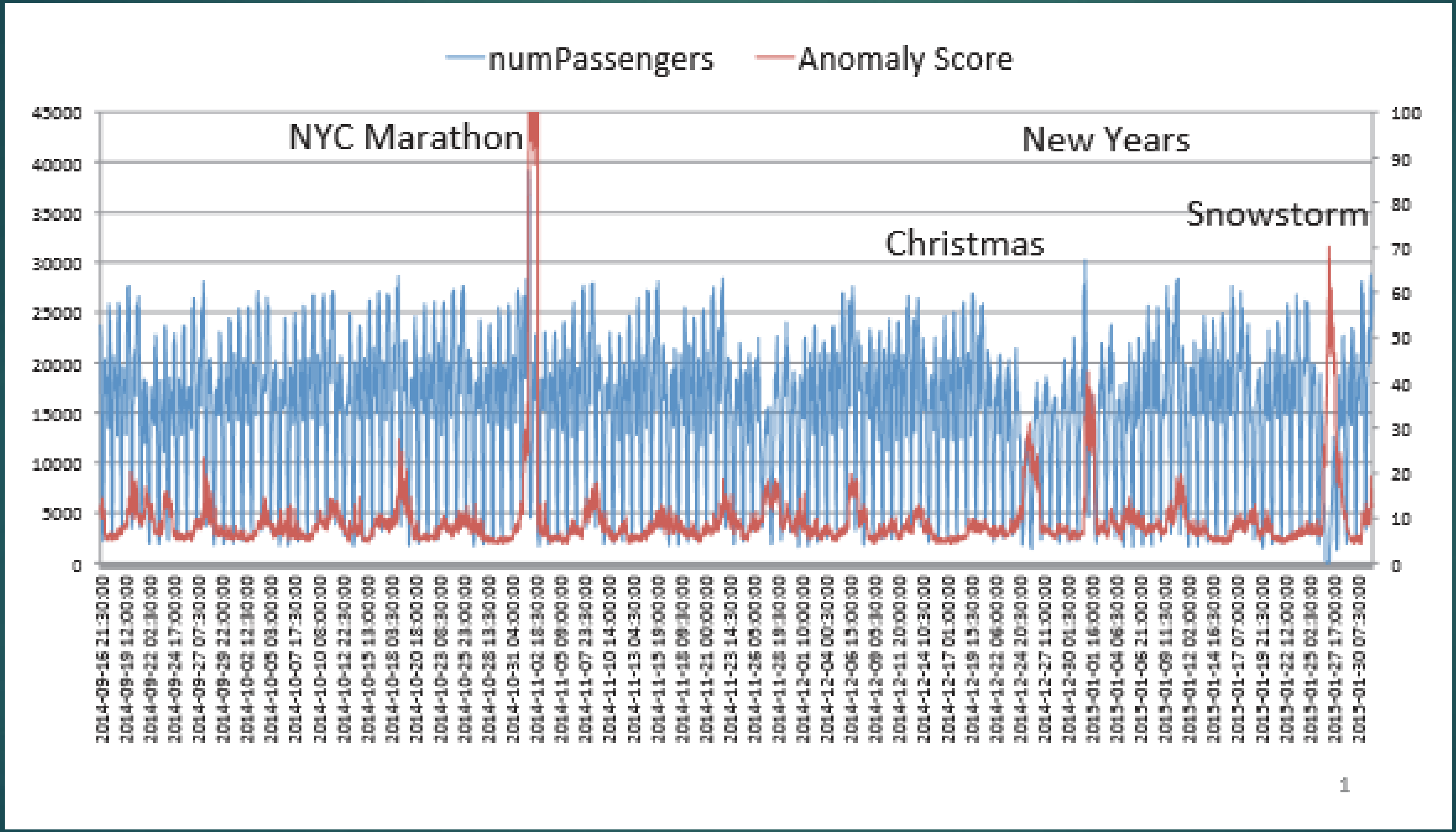
2.5 Ensemble of RRCTs

- Those theorems can be expanded to include in ensemble case:
  - The probability of choosing a random cut that splits S is exactly same as the conditional probability of choosing a random cut that splits $S \cup \{p\}$ conditioned on not isolating p from all points of S
  - Insertion, deletion also follow same definitions

2.6 Viability

- New York taxi ridership
- Shingling data
  - Each time stamp as a different feature for a window with stride of 1
  - It can capture typical shape, any departure can be interpreted as anomaly
- Data collected for 7 months
- Data is 1-dimensional but by shingling, they include current day and last day, so 48 dimensions
- Special days as anomaly
- Accuracy 96 and AUC 0.9

2.6 Viability



The end
Thank you