Cloud Service Modelling

For Service Diagnostic and Anomaly Detection

Viveka Kulharia (MSR)
Arushi Jain (MSR)
Sundararajan S. (MSR)

Why Cloud Service Modelling?

- Provide automated, comprehensive, near-real time insights into the cloud service based on high dimensional counter time-series data (measures performance of the service).
- Currently, most of the monitoring tools are heuristics based which fails to capture joint relation among large set of time-series, leading to false positives.
- Provides hierarchical monitoring capturing relation among high dimensional time to identify overall health state of service.

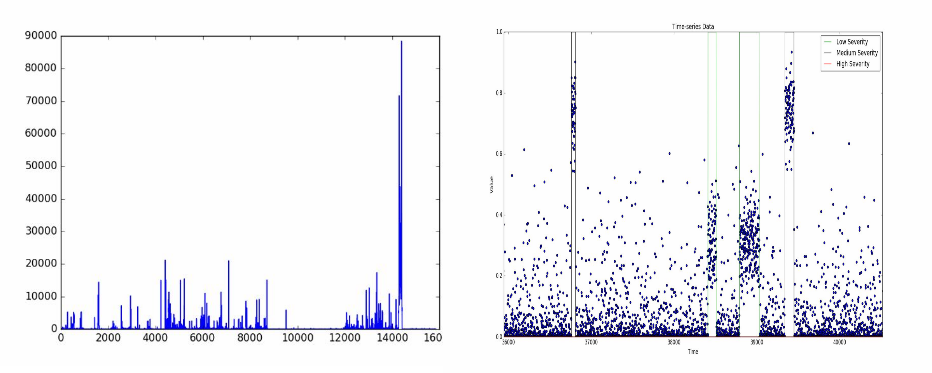
Challenges

- Dearth of labelled data
- High dimensional time-series
- Real-time anomaly detection
- Unknown behavior of counter time-series

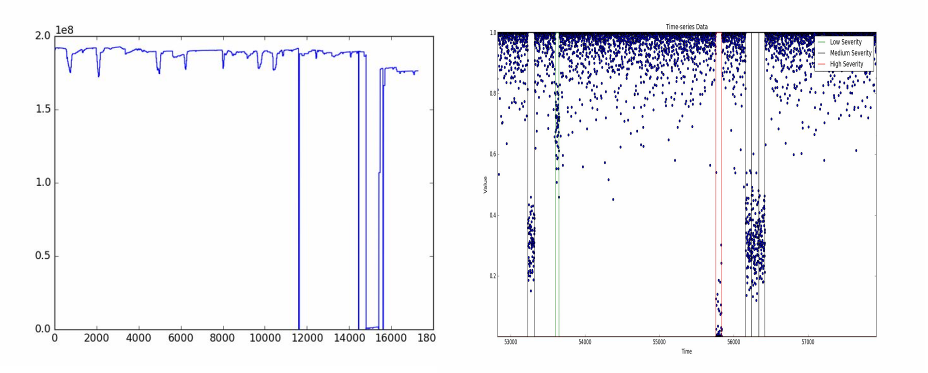
Solution

- Probabilistic Cloud Service Modelling - generates synthetic data, can be used to evaluate any anomaly detection algorithm.
- Segmentation of individual time-series based on anomalous behavior.
- Labelling segments based on severity of the issue
- Joint Modelling to identify the overall health state of service

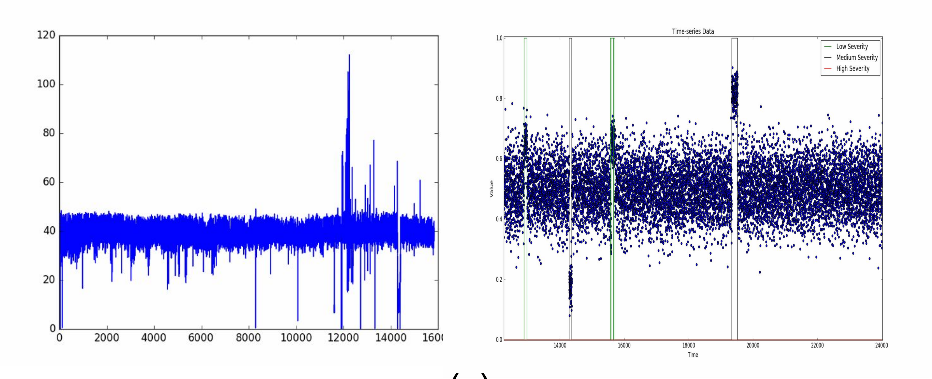
Real Vs Generated time-series



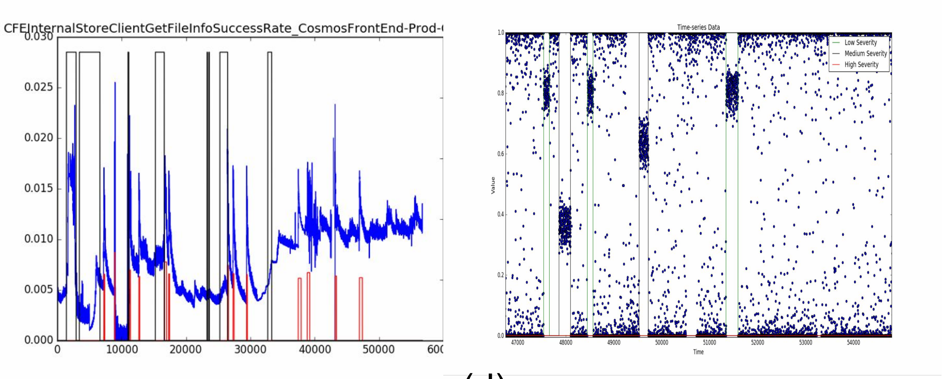
(a)



(b)



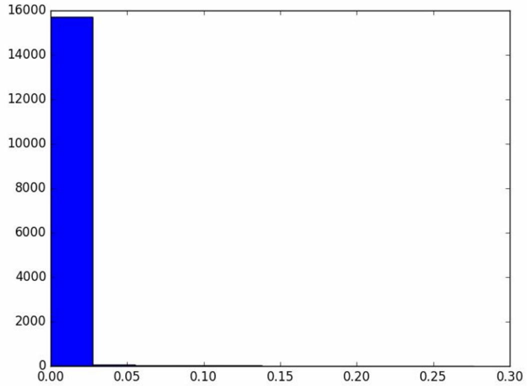
(c)



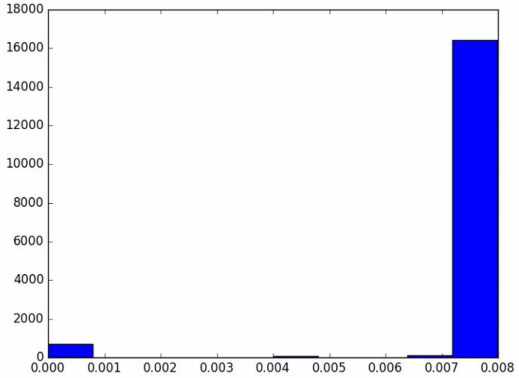
(d)

Histograms

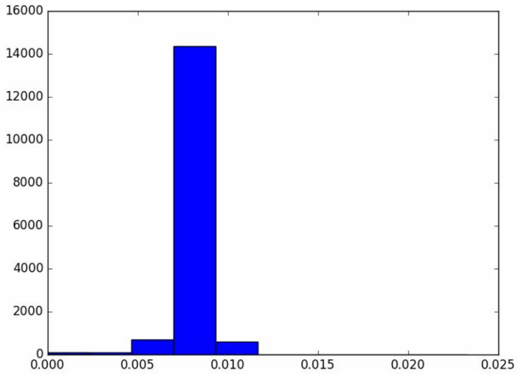
(a) Left Exponential



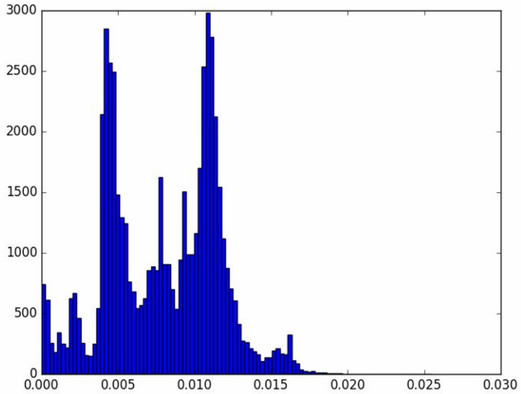
(b) Right Exponential



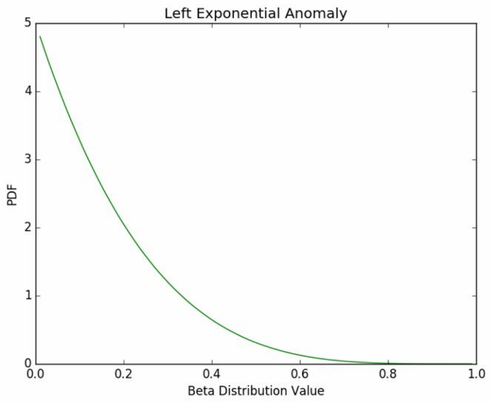
(c) Inverted U shape



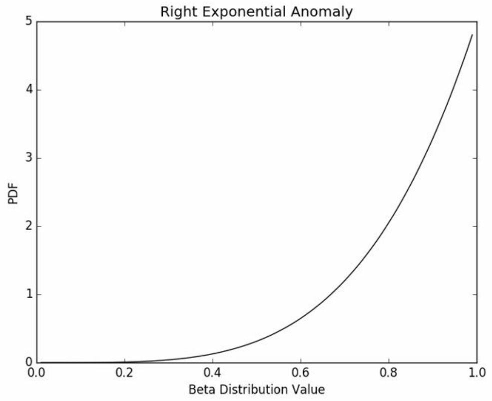
(d) U shape

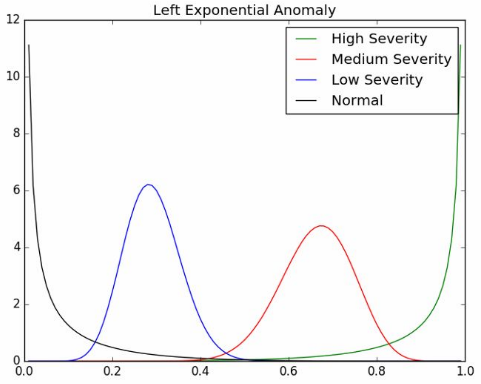


Probabilistic Modeling

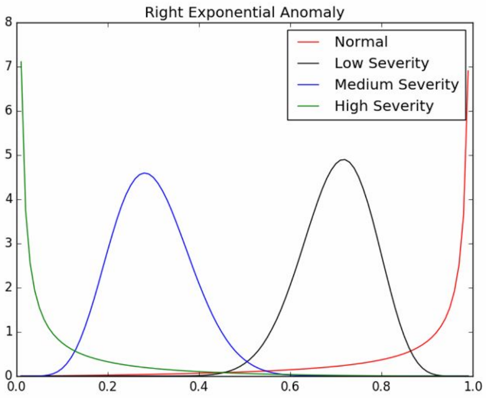


Distribution of time series

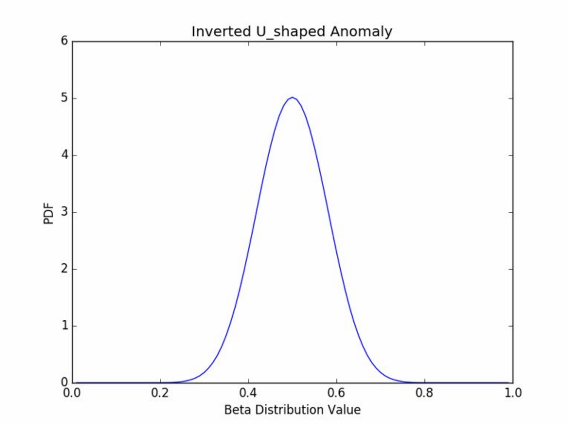




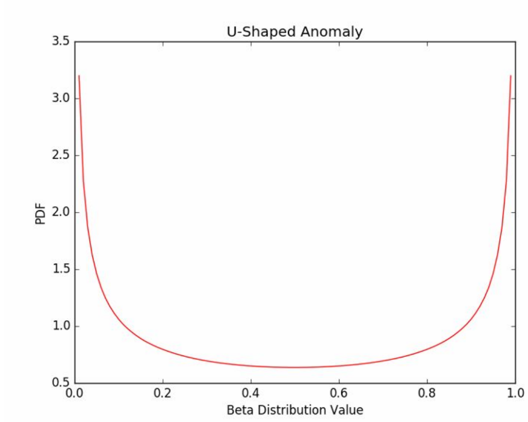
Mixture of beta distributions with anomaly severity

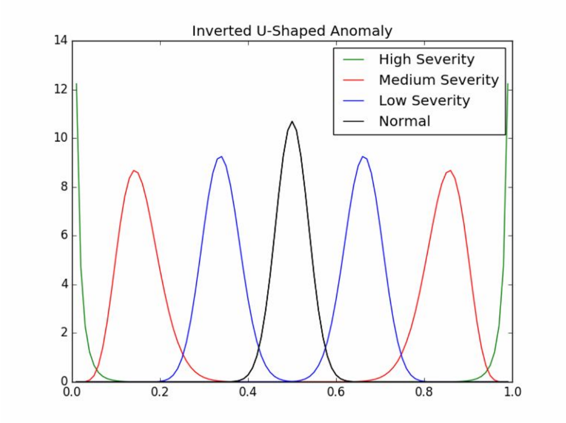


Probabilistic Modeling

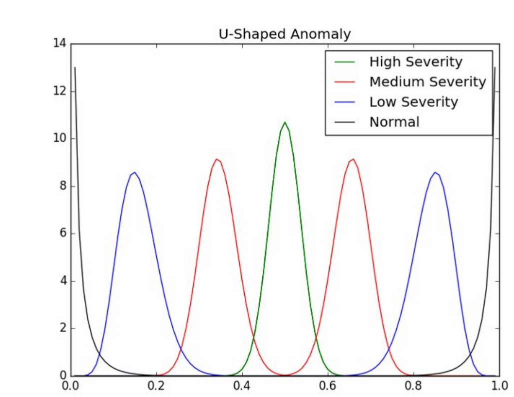


Distribution of time series





Mixture of beta distributions with anomaly severity



Segmentation of individual time-series

Binary Segmentation: $O(nlog(n))$ method that works by recursively finding the change-points in a given time segment

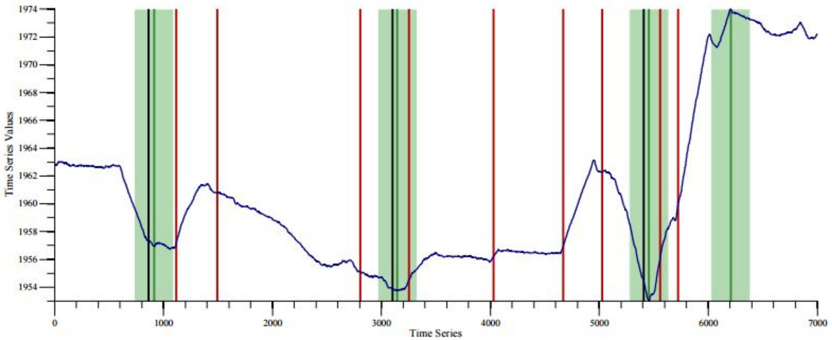
1. Linear Binary Segmentation
2. Optimal Linear Binary Segmentation
3. Mid Point Tester

Optimal Segmentation: $O(n^2)$ dynamic programming approach method that optimizes the objective $\min \Sigma_{i=1}^{m+1}[C(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m)$

PELT: improves the running time of the above method by reducing the search space. Under certain assumptions this method given O(n) time

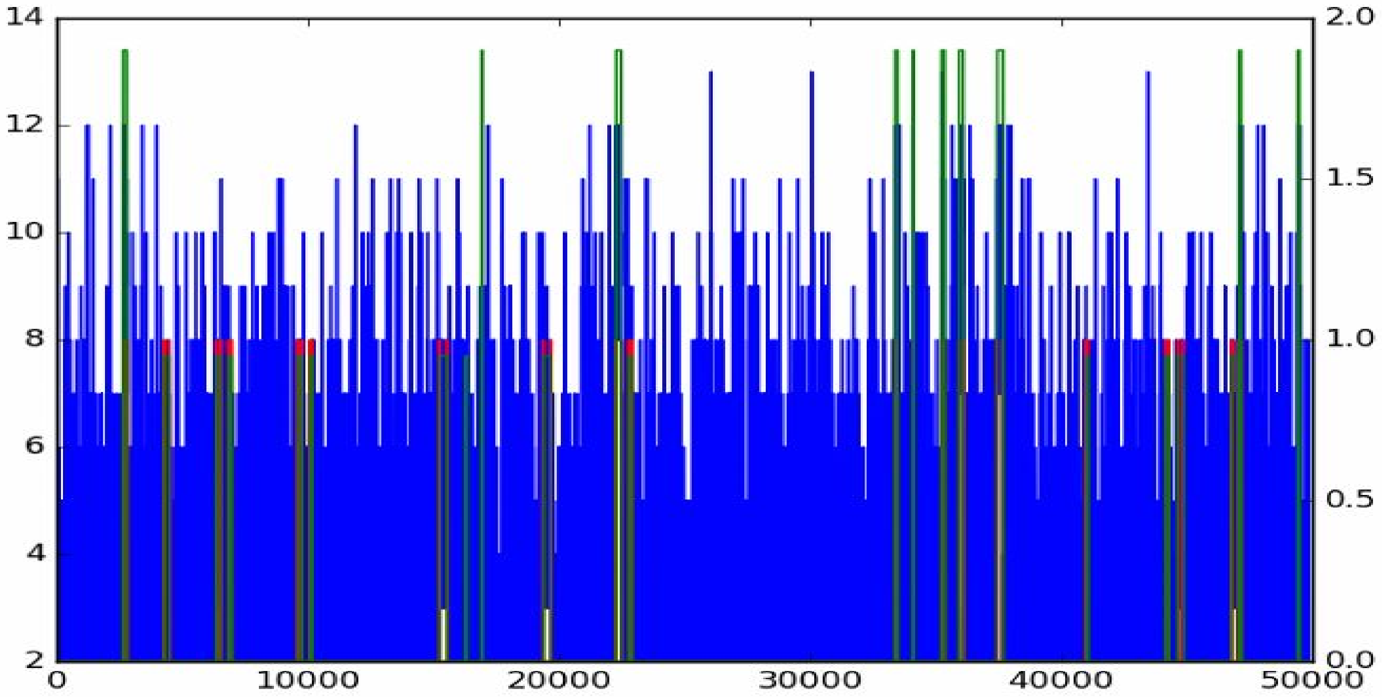
Segmentation Evaluation

True Positive (TP): Predicted segmentation point lying in segmentation zone (SZ) surrounding true point.
False Positive (FP): Predicted segmentation point lying inside SZ when there is already one more or when predicted point is outside SZ.
False Negative (FN): True segmentation point not covered by any predicted point within SZ
True Negative (TN): Any point which is none of TP, FP or FN.



Green vertical lines denote actual segment point and Green area denotes their segmentation zone.
Black lines are predicted segment points considered as TP while Red lines are predicted segment points considered as FP.
Source: “Novel Criteria to Measure Performance of Time Series Segmentation Techniques”

Synthetic Segmentation



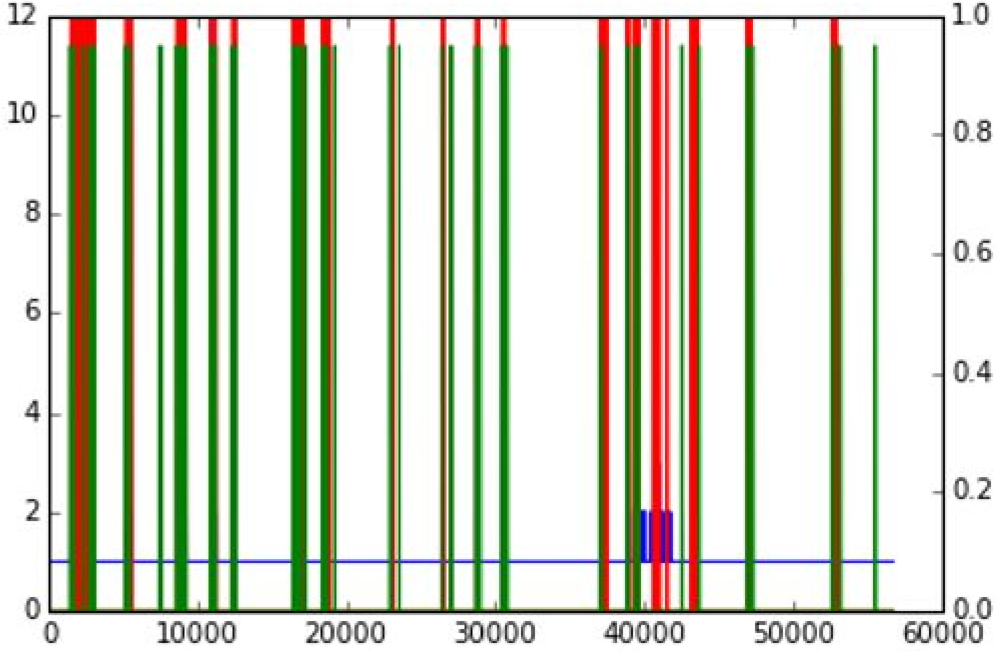
Green lines are actual segment points (height: 1 or 2) while red is the predicted segment points.

Evaluation of Segmentation

Counter	Precision	Recall	F1
1	92.30%	90%	91.13%
2	80.55%	85.29%	82.85%
3	100%	94.28%	97.05%
4	100%	94.66%	97.26%
5	98.48%	86.66%	92.19%

*On dataset with results on 5 counters

Real time-series Segmentation



Green lines are actual segment points while red is the predicted segment points.
Blue is the discretized time series.

Evaluation on Real time-series

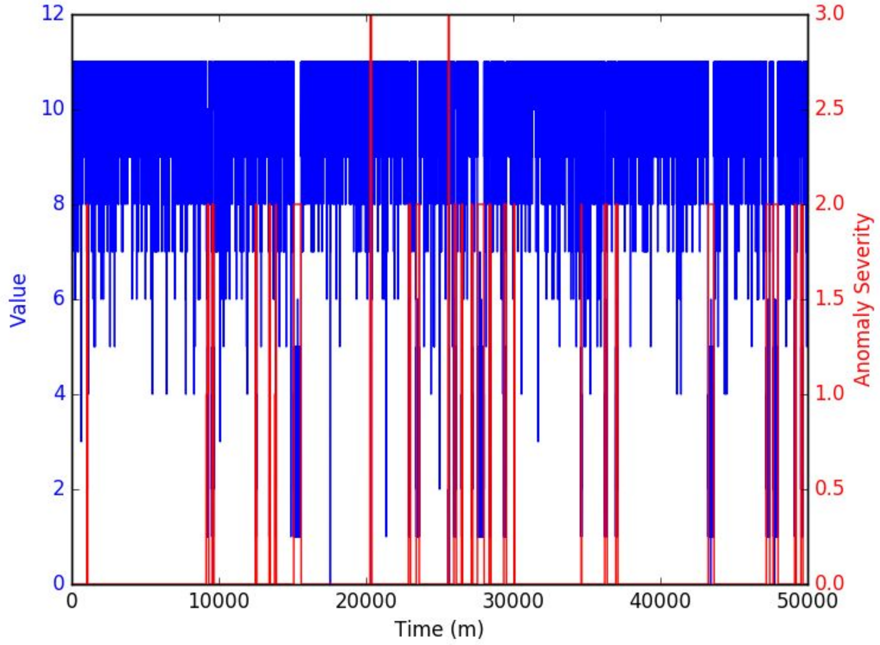
Counter	Precision	Recall	F1
JobsSuccessRate	45.71%	72.72%	56.14%
ClientEnumDirectorySuccessRate	42.10%	72.72%	53.33%
ServiceEnumDirectorySuccessRate	32.25%	86.95%	47.05%
RequestsCurrent	37.03%	50%	42.55%
QueryQueueSuccessRate	25.71%	75%	38.29%

*For Cy2 dataset

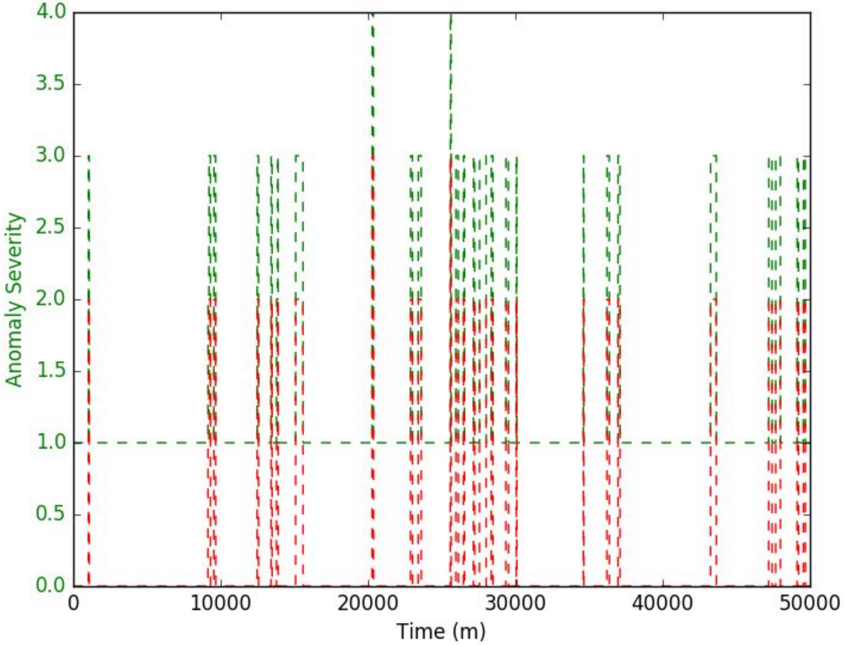
Labelling

- The segments obtained on individual time-series were labelled according to severity of the segment anomalous behavior.
- Labels: Normal state, Low Severity, Medium Severity and High Severity
- Approach
  - Alpha-Beta parameter estimation of each segment in time-series
  - Labelling classifier which classifies into 4 states

Synthetic Labeling



Blue lines are the time series data in which one has to detect the anomalies and red lines are the actual anomaly severities (0: no severity, 3: high severity)



Red lines are actual anomaly severity (0: no severity, 3: high severity) while green lines (shifted up by a unit for clarity) are predicted anomaly severities.

Labelling metrics

- 4x4 Confusion matrix is created using the predicted label and actual label for each point of the time series.
- Precision, recall and F1 score is computed for each of the 4 labels.
- The precision, recall and F1 score reported on next slide is the average for the respective scores across each labels.

Evaluation on Labelling

Counter	Precision	Recall	F1
1	99.72%	95.89%	97.72%
2	99.37%	91.89%	95.24%
3	98.74%	82.04%	87.42%
4	99.45%	96.11%	97.71%
5	96.86%	58.19%	91.16%

*On dataset with results on 5 counters

Joint Segmentation and Labeling

- Grouping the counter time-series according to anomalous behavior to classify the overall health state of the service.
- Health state of service: Grouping counters based on anomaly.
    - Normal State (No anomalous behavior)
    - Known Issue Type (Low, Medium, High Severity): Relationship among anomalous counters is known.
    - Unknown Issue Type (Low, Medium, High Severity): Identifying the latent relationship among counter time-series.
- Viterbi based Joint modelling approach

Notation

- KIT denotes Known Issue Type
- UKIT denotes Unknown Issue Type
- Is$\in \{N, KIT_1, KIT_2, \cdots, KIT_n, UKIT\}$:Where Is denotes type of issue occurring in service
- He$\in \{N, KIT_H, KIT_M, KIT_L, UKIT_H, UKIT_M, UKIT_L\}$
  :Where He denotes overall health state of the service

Evaluation on Joint Segmentation and Labeling

Results on labelling of time-series at overall health state level

Anomalous Type	Precision	Recall	F1 score
Normal	99.5%	99.2%	99.4%
UKIT-Low severity	88.5%	100%	93.9%
UKIT-Medium severity	92.1%	91.7%	91.9%
UKIT-High severity	94.3%	84.8%	89.3%
KIT-Low severity	99.2%	97.9%	98.6%
KIT-Medium severity	100%	97.3%	98.6%
KIT-High severity	100%	40.8%	58.0%

Issue Type Accuracy

For KIT, 4 issue types were taken, where each type comprises of set of counter involved in anomalous behavior. Below table summarizes the performance metric for accuracy of predicting issue type class at overall health state level.

Issue Type	Precision	Recall	F1 score
KIT- Issue Type1	100%	98.8%	99.4%
KIT- Issue Type2	100%	88.8%	94.1%
KIT- Issue Type3	100%	98.5%	99.3%
KIT- Issue Type4	100%	98.3%	99.1%
Normal, UKIT	99.7%	99.7%	99.9%