DETECTION OF ANOMALOUS EVENTS ON INTERNET PATHS

Presented By:
Fida Gillani
Research Assistant
CyberDNA, UNCC
Outline

- Introduction
- Motivation
- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection
- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
  - Comparative Accuracy Evaluation with Existing Techniques
- Conclusion and Summary of Contribution
- References
Outline

- **Introduction**
  - Motivation
  - Existing Work
    - Contemporary Anomalous Event Detectors
      - Most existing algorithms cannot be adopted
      - Algorithms adopted for Path Event Detection
  - Research Contributions
    - Dataset Selection
    - Dataset Labeling
    - New Event Detection Algorithm
    - Comparative Accuracy Evaluation with Existing Techniques
- **Conclusion and Summary of Contribution**
- **References**
Detection of Anomalous Events on Internet Paths

- **Internet Path**: A series of links used to transfer data from source to destination
Introduction

Detection of Anomalous Events on Internet Paths

- **Anomalous**: Any abnormal behavior, such as:
  - Security Threats (Port Scans, Worms, DOS attacks etc)
  - Equipment Failures (End host Failure, Link outage)
  - Uncharacteristic Usage (Flash Crowds, High volume flows)
  - Uncharacteristic Behavior (Misconfigurations, Fluttering in routes)
Detection of Anomalous Events on Internet Paths

- **Event:** A significant change in the state of a path, that persists for some duration
- **So an event is characterized by three parameters:**
  - **State of a Path:**
    - **Throughput Achievable:** Maximum bandwidth achieved by an application
    - **Available Bandwidth:** Difference between total bandwidth and throughput achieved by all applications
Introduction

Detection of Anomalous Events on Internet Paths

- **Event**: A significant change in the state of a path, that persists for some duration

- So an **event** is characterized by three parameters:
  - State of a Path:
    - **Throughput Achievable**: Maximum bandwidth achieved by an application
    - **Available Bandwidth**: Difference between total bandwidth and throughput achieved by all applications
**Introduction**

**Detection of Anomalous Events on Internet Paths**

- **Event**: A significant change in the state of a path, that persists for some duration

- So an *event* is characterized by three parameters:
  - **State of a Path**:
    - **Throughput Achievable**: Maximum bandwidth achieved by an application
    - **Available Bandwidth**: Difference between total bandwidth and throughput achieved by all applications
  - Magnitude of the change in the state of the Internet Path is significant
  - Persistence of significant change for some duration
In this research, we aim to design, implement and evaluate an algorithm that can detect Internet path events in an accurate and timely manner.
In this research, we aim to design, implement and evaluate an algorithm that can detect Internet path events in an accurate and timely manner.

- **Step 1**: Load is balanced properly by Router A
- **Step 2**: Link is down in Path 2
- **Step 3**: Load is shifted to Path 1 by Router A
- **Step 4**: If problem remains for some duration
- **Step 5**: Two alarms are raised, for Path 2 “No available bandwidth” and for Path 1 “Minimum Available Bandwidth”.

Anomaly Detector: At Step 4 Two alarms are raised for Path 1 & 2.
Outline

- Introduction

- Motivation

- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection

- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
  - Comparative Accuracy Evaluation with Existing Techniques

- Conclusion and Summary of Contribution

- References
General Motivation

- Solution to this problem will be helpful in:
  - Network Operations
    - E.g.: Identifying changes in the network path, quantifying magnitude of the changes, providing alerts and determining whether fault lies with the network path or the applications
  - Planning
    - To improve performance today and in the future by providing historical information on growth, incremental and sudden changes
  - Maintaining Academic & Research Networks

- Performance evaluation of existing techniques for Internet Path Event Detection does not exist

- Feedback to TCP congestion
Specific Motivation

- LHC computing required a worldwide grid
- SLAC undertook a project
  - Measurement and Evaluation of state of Internet path
  - Help critical grid application of high energy physics experiments
- Idea was:
  - To adapt current techniques for fault diagnosis on Internet path
  - To develop new techniques
- Main emphasis was to benefit grid applications, especially for Worldwide LHC Computing Grid (WLCG)
Outline

- Introduction
- Motivation

Existing Work

- Contemporary Anomalous Event Detectors
  - Most existing algorithms cannot be adopted
  - Algorithms adopted for Path Event Detection

Research Contributions

- Dataset Selection
- Dataset Labeling
- New Event Detection Algorithm
- Comparative Accuracy Evaluation with Existing Techniques

Conclusion and Summary of Contribution

References
## Contemporary Anomalous Event Detectors

<table>
<thead>
<tr>
<th></th>
<th>Anomalous Event Detectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Plateau Algorithm by Stanford Linear Accelerator [1]</td>
</tr>
<tr>
<td>2</td>
<td>Holt-Winter Forecasting based Algorithm [4]</td>
</tr>
<tr>
<td>3</td>
<td>Kalman Filter based Algorithm by Intel Research [2]</td>
</tr>
<tr>
<td>4</td>
<td>Adaptive Fault Detection Algorithm [3]</td>
</tr>
<tr>
<td>5</td>
<td>Maximum Entropy Method [5]</td>
</tr>
<tr>
<td>6</td>
<td>Principal Component Analysis based Algorithm [6]</td>
</tr>
<tr>
<td>7</td>
<td>Rate Limiting Algorithm [7]</td>
</tr>
<tr>
<td>8</td>
<td>Threshold Random Walk Algorithm [8]</td>
</tr>
<tr>
<td>9</td>
<td>TRW with Credit based Limiting Algorithm [9]</td>
</tr>
<tr>
<td>10</td>
<td>Packet Header Anomaly Detection [10]</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Motivation

**Existing Work**

- Contemporary Anomalous Event Detectors
  - Most existing algorithms cannot be adopted
  - Algorithms adopted for Path Event Detection

**Research Contributions**

- Dataset Selection
- Dataset Labeling
- New Event Detection Algorithm
- Comparative Accuracy Evaluation with Existing Techniques

**Conclusion and Summary of Contribution**

**References**
Most existing algorithms cannot be adopted

<table>
<thead>
<tr>
<th>Sr.#</th>
<th>Anomalous Event Detector Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maximum Entropy Method based Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Principal Component Analysis based Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Rate Limiting Algorithm</td>
</tr>
<tr>
<td>4</td>
<td>Threshold Random Walk Algorithm</td>
</tr>
<tr>
<td>5</td>
<td>Packet Header Anomaly Detection Algorithm</td>
</tr>
<tr>
<td>6</td>
<td>TRW with Credit based Limiting Algorithm</td>
</tr>
</tbody>
</table>

These algorithms flag anomalous events using TCP/IP header fields
Most existing algorithms cannot be adopted

<table>
<thead>
<tr>
<th>Sr.#</th>
<th>Anomalous Event Detector Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maximum Entropy Method based Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Principal Component Analysis based Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Rate Limiting Algorithm</td>
</tr>
<tr>
<td>4</td>
<td>Threshold Random Walk Algorithm</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

- Distributions of origin destination (OD) flows
- Transport ports

These algorithms flag anomalous events using TCP/IP header fields
Most existing algorithms cannot be adopted

<table>
<thead>
<tr>
<th>Sr. #</th>
<th>Anomalous Event Detector Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maximum Entropy Method based Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Principal Component Analysis based Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Rate Limiting Algorithm</td>
</tr>
<tr>
<td>4</td>
<td>Threshold Random Walk Algorithm</td>
</tr>
<tr>
<td>5</td>
<td>Packet Header Anomaly Detection Algorithm</td>
</tr>
<tr>
<td>6</td>
<td>TRW with Credit based Limiting Algorithm</td>
</tr>
</tbody>
</table>

Use frequency of connections between hosts.

These algorithms flag anomalous events using TCP/IP header fields.
Most existing algorithms cannot be adopted

<table>
<thead>
<tr>
<th>Sr.#</th>
<th>Anomalous Event Detector Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maximum Entropy Method based Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Principal Component Analysis based Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Rate Limiting Algorithm</td>
</tr>
<tr>
<td>4</td>
<td>Threshold Random Walk Algorithm</td>
</tr>
<tr>
<td>5</td>
<td>Packet Header Anomaly Detection Algorithm</td>
</tr>
<tr>
<td>6</td>
<td>TRW with Credit based Limiting Algorithm</td>
</tr>
</tbody>
</table>

- Uses all 33 fields of the Ethernet, IP, TCP/UDP headers to calculate an anomaly score

These algorithms flag anomalous events using TCP/IP header fields
Most existing algorithms cannot be adopted

<table>
<thead>
<tr>
<th>Sr.#</th>
<th>Anomalous Event Detector Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maximum Entropy Method based Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Principal Component Analysis based Algorithm</td>
</tr>
<tr>
<td>3</td>
<td>Rate Limiting Algorithm</td>
</tr>
<tr>
<td>4</td>
<td>Threshold Random Walk Algorithm</td>
</tr>
<tr>
<td>5</td>
<td>Packet Header Anomaly Detection Algorithm</td>
</tr>
<tr>
<td>6</td>
<td>TRW with Credit based Limiting Algorithm</td>
</tr>
</tbody>
</table>

These algorithms flag anomalous events using TCP/IP header fields.

**These traffic features cannot characterize the state of an Internet Path**
Outline

- Introduction
- Motivation

**Existing Work**
- Contemporary Anomalous Event Detectors
  - Most existing algorithms cannot be adopted
  - Algorithms adopted for Path Event Detection

- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
  - Comparative Accuracy Evaluation with Existing Techniques

- Conclusion and Summary of Contribution

- References
Candidate State Parameters

- Round Trip Time
- Packet Loss
- One Way Delay
- Network Path
- Available Bandwidth
- Achievable Throughput
For accurate anomaly detection, an anomaly detectors should consider the inherent statistical nature of the data, but all these techniques were implying some assumed statistical behavior.
Outline

- Introduction
- Motivation
- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection

- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
  - Comparative Accuracy Evaluation with Existing Techniques

- Conclusion and Summary of Contribution
- References
Dataset Selection

- We use performance measurements from the Internet End-to-end Performance Monitoring Bandwidth (IEPM-BW) project [11]

- IEPM-BW targets high performance network links used worldwide by Academic & Research Applications
Dataset Selection

- Three measurements tools are deployed on IEPM-BW
  - Iperf [12], Thrulay [13] and PathChirp [14]
  - Iperf & Thrulay provide throughput achievable and PathChirp provides available bandwidth (all units in Mbps)

- **Iperf**
  - Sends bulk of TCP streams
  - Estimates throughput using total data transferred over a time interval

- **Thrulay**
  - Same as Iperf
  - Provides one way delay by sending UDP packets

- **PathChirp**
  - Uses concept of self-induced congestion
  - Sends exponentially spaced packets in the form of chirps/packet trains
  - Uses packet inter arrival delay to estimate available bandwidth
Dataset Selection

- The dataset selection criteria were:
  - Measurements from all three performance tools Iperf, Thrulay and Pathchirp should be available;
  - Should have minimum downtime;
  - Should span international boundaries; and
  - Should contain diverse types of traffic (e.g., FTP, HTTP, multimedia, encrypted traffic, etc.)

- We selected 21 datasets from the IEPM-BW database
## Dataset Selection: From SLAC to

<table>
<thead>
<tr>
<th>Sr.#</th>
<th>Path Name</th>
<th>Sr.#</th>
<th>Path Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>San Diego Supercomputing Center (SDSC) USA</td>
<td>12</td>
<td>Node1.Ornal.Gov</td>
</tr>
<tr>
<td>2</td>
<td>European Organization for Nuclear Research (CERN) Geneva, Switzerland</td>
<td>13</td>
<td>Bnl_Edu</td>
</tr>
<tr>
<td>3</td>
<td>Forschungszentrum Karlsruhe (FZK) Germany</td>
<td>14</td>
<td>Node1_mcs_anl_go</td>
</tr>
<tr>
<td>4</td>
<td>Deutsches Elektronen-Synchrotron (DESY) Germany</td>
<td>15</td>
<td>Bnl_org</td>
</tr>
<tr>
<td>5</td>
<td>Oak Ridge National Laboratory (ORNL) USA</td>
<td>16</td>
<td>Caltec_Edu</td>
</tr>
<tr>
<td>6</td>
<td>University of Toronto (UTORONTO) Canada</td>
<td>17</td>
<td>Caltec_Org</td>
</tr>
<tr>
<td>7</td>
<td>Stanford University USA</td>
<td>18</td>
<td>Nod1_dl_ac_ul</td>
</tr>
<tr>
<td>8</td>
<td>Node1_Switch_Ch</td>
<td>19</td>
<td>Node1_pd_infn_i</td>
</tr>
<tr>
<td>9</td>
<td>Node1_Nslabs_ufl</td>
<td>20</td>
<td>Node1_cesnet_cs</td>
</tr>
<tr>
<td>10</td>
<td>Node1_binp_snk_su</td>
<td>21</td>
<td>Node2_Nslabs_Ufl</td>
</tr>
<tr>
<td>11</td>
<td>Node1_dl_ac_ul</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dataset Selection: Discrepancies in Measurements

- All tools should track each other on the same path and during the same time window
- But we have found clear discrepancies in PathChirp and Iperf/Thrulay measurements
- Consider the following example on a link from SLAC to SDSC:

<table>
<thead>
<tr>
<th>Date</th>
<th>PathChirp</th>
<th>Iperf</th>
<th>Thrulay</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/23/2006</td>
<td>1041.63</td>
<td>429</td>
<td>518.8</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1142.08</td>
<td>429</td>
<td>503.9</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1110.6</td>
<td>430</td>
<td>497.1</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1167.46</td>
<td>431.1</td>
<td>350.6</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>965.42</td>
<td>400</td>
<td>430.6</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1076.73</td>
<td>401</td>
<td>508</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1115.88</td>
<td>420</td>
<td>515.5</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1056.18</td>
<td>429</td>
<td>452.8</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1209.1</td>
<td>429</td>
<td>418.3</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1226.46</td>
<td>431.1</td>
<td>469.1</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1119.184</td>
<td>249.1</td>
<td>431.7</td>
</tr>
<tr>
<td>3/23/2006</td>
<td>1041.63</td>
<td>429</td>
<td>518.8</td>
</tr>
</tbody>
</table>
Dataset Selection: Discrepancies in Measurements

![Graph showing available bandwidth (Mbps) over time and frequency distribution.](image)
Dataset Selection: Discrepancies in Measurements

Pathchirp

Thru1ay
Thrunlay shows reverse exponential distribution, showing some kind of a barrier
- The discrepancy is due to the use of TCP Reno congestion stack which is used by both Iperf & Thrunlay
- TCP Reno works well for low speed links but fails for high speed links, already discussed in following paper [15]
Our findings suggest that the most appropriate tool for Internet path monitoring is PathChirp and the corresponding path measurement metric is Available Bandwidth
Outline

- Introduction
- Motivation
- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection

Research Contributions

- Dataset Selection
- Dataset Labeling
- New Event Detection Algorithm
- Comparative Accuracy Evaluation with Existing Techniques

- Conclusion and Summary of Contribution
- References
Dataset Labeling

- Event definition
  - “Significant change in state that persists for some time”

- This definition is vague because it does not quantify the:
  - Magnitude of change
  - Duration of change
Traditionally, mean bandwidth values have been used for event detection.

But the data contains random spikes.

So the question is:

– Should we also consider the variance in the bandwidth estimates as a qualitative measure of change?
The question is:
- Should we also consider the variance in the bandwidth estimates as a qualitative measure of change?

To address this question, we filter out the anomalous bandwidth values using a median filter [16]

The filtered data is then evaluated to quantify the mean and variance of the baseline behavior
Magnitude of Change: Median Filter

Low-pass median filtering of bandwidth measurements to extract baseline behavior; the time series is annotated to show how median filtering results in removal of sustained anomalies and spurious measurements.
We assume these 6% values as anomalous
  – If this assumption is true these anomalous values should possess the property of persistence

We plot a subset (mixture of values with different deviations from the mean) of the dataset
  – Red line is baseline mean
All values with a difference greater than .5 mean show consistency.

The only difference between two states is in the mean not in the variance.
To statistically substantiate these findings, we inspect the temporal dependency within data

- We perform conditional entropy based Markov chains analysis
  - Conditional Entropy
    - Entropy is a measure of uncertainty in a random source
    - Conditional Entropy is the amount of uncertainty left given an event has occurred
  - Markov Chains
    - Information available to predict the next state provided the system is in some state
Magnitude of Change: Markov Chain Analysis

- **Intuition behind this analysis:**
  - If bandwidth measurements are produced from a benign source then they should possess significant temporal dependence.
  - If benign data is divided into states then the amount of information present in one state about the next state is high.

| 10 | 02 | 12 | 14 | 09 | 16 | 08 | 06 | 14 | 15 | 25 | 21 | 20 | 14 | 13 | 10 |

- 0 to 9
- 10 to 19
- 20 to 30
Intuition behind this analysis:

- If bandwidth measurements are produced from a benign source then they should possess significant temporal dependence.
- If benign data is divided into states then the amount of information present in one state about the next state is high and this is *markov property*.

\[
\Pr\{X_n = j \mid X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0\} = \Pr\{X_n = j \mid X_{n-1} = i\}
\]

- We divide data into states to perform Markov Chain analysis.
Magnitude of Change: Markov Chain Analysis

- We check the patterns of temporal dependency on each subset

\[ H^{(1)} = - \sum_{i \in \psi^{(1)}} \pi_i^{(1)} \sum_{j \in \psi^{(1)}} \left( p_{X_{n=j}|X_{n-1}=i}^{(1)} \right) \log_2 \left( p_{X_{n=j}|X_{i-1}=i}^{(1)} \right) \]

We have found two different temporal dependency patterns around

[\gamma \leq 0.5 \quad \kappa \geq 1.5]
Magnitude of Change: SLAC to BNL

(a) BNL as seen from SLAC
Magnitude of Change: SLAC to DESY

(b) DESY as seen from SLAC
## Duration of Change?

<table>
<thead>
<tr>
<th>Path Name</th>
<th>3 Hrs Event Window</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTORONTO</td>
<td>38</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
</tr>
<tr>
<td>DESY</td>
<td>31</td>
</tr>
<tr>
<td>SDSC</td>
<td>6</td>
</tr>
<tr>
<td>NODE1_FZK</td>
<td>17</td>
</tr>
<tr>
<td>NODE1_TRIUMF</td>
<td>3</td>
</tr>
<tr>
<td>NODE1_NSLABS_EFL</td>
<td>4</td>
</tr>
<tr>
<td>UTORONTO</td>
<td>38</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
</tr>
</tbody>
</table>
Dataset Labeling: Conclusion

- **Magnitude of change**
  - Deviation to the left of the mean: $\gamma \leq 0.5$
  - Deviation to the right of the mean: $\kappa \geq 1.5$

- **Duration of change**
  - 6 value window is suitable
Outline

- Introduction
- Motivation
- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection
- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
    - Comparative Accuracy Evaluation with Existing Techniques
- Conclusion and Summary of Contribution
- References
Why do we need a new algorithm?

Existing anomaly detectors do not consider the inherent nature of data
We use a decision-theoretic approach to devise a new anomalous event detector.

- It considers two choices called Hypothesis:
  - Hypothesis $H_1$: traffic observed in the last $T$ seconds is anomalous
  - Hypothesis $H_0$: traffic observed in the last $T$ seconds is NOT anomalous

- $H_1$ Detection Hypothesis & $H_0$ Null Hypothesis
New Event Detection Algorithm: Decision Theoretic Approach

- In our case let $R_i$ be the bandwidth measurements
  - $H_0 : R_i = n$
  - $H_1 : R_i = n + m_i$

- Here $n$ represents the baseline distribution & $m_i$ represents the noise.

- We observe a multiple-distribution behavior using chi-square test and graphical analysis
  - Gaussian, Weibull, logistic and beta general distributions (multiple-Distribution behavior) are observed, one distribution in one dataset.

- All these distributions are used in decision-theoretic approach for anomalous event detection.
### Decision-Theoretic Approach

<table>
<thead>
<tr>
<th>$R_i$</th>
<th>10</th>
<th>02</th>
<th>12</th>
<th>14</th>
<th>09</th>
<th>16</th>
<th>08</th>
<th>06</th>
<th>14</th>
<th>15</th>
<th>25</th>
<th>21</th>
<th>20</th>
<th>14</th>
<th>13</th>
<th>10</th>
</tr>
</thead>
</table>
$H_0 : R_i = n$

<table>
<thead>
<tr>
<th>$R_i$</th>
<th>10</th>
<th>02</th>
<th>12</th>
<th>14</th>
<th>09</th>
<th>16</th>
<th>08</th>
<th>06</th>
<th>14</th>
<th>15</th>
<th>25</th>
<th>21</th>
<th>20</th>
<th>14</th>
<th>13</th>
<th>10</th>
</tr>
</thead>
</table>

Decision-Theoretic Approach
$H_0 : R_i = n$

R_i | 10 | 02 | 12 | 14 | 09 | 16 | 08 | 06 | 14 | 15 | 25 | 21 | 20 | 14 | 13 | 10

Gaussian

Weibull

(a) Logistic & Beta Journal
Decision-Theoretic Approach

\[ H_0 : R_i = n \]

<table>
<thead>
<tr>
<th>R_i</th>
<th>10</th>
<th>02</th>
<th>12</th>
<th>14</th>
<th>09</th>
<th>16</th>
<th>08</th>
<th>06</th>
<th>14</th>
<th>15</th>
<th>25</th>
<th>21</th>
<th>20</th>
<th>14</th>
<th>13</th>
<th>10</th>
</tr>
</thead>
</table>

Figure 1: Weibull Density Function

(a) Gaussian

(b) Weibull

Logistic & Beta Journal
Decision-Theoretic Approach

$H_0 : R_i = n$

<table>
<thead>
<tr>
<th>$R_i$</th>
<th>10</th>
<th>02</th>
<th>12</th>
<th>14</th>
<th>09</th>
<th>16</th>
<th>08</th>
<th>06</th>
<th>14</th>
<th>15</th>
<th>25</th>
<th>21</th>
<th>20</th>
<th>14</th>
<th>13</th>
<th>10</th>
</tr>
</thead>
</table>

Gaussian

Weibull

$\text{density}$

$\text{density}$

(a)

(b)

Figure 1

Weibull Density Function

- Exponential, $m = 1.0$
- $m = 1.5$
- $m = 3.51$
- $m = 8.5$

$n + m_i$
**Decision-Theoretic Approach**

\[ H_0 : R_i = n \]

\[
\begin{array}{cccccccccccccccc}
R_i & 10 & 02 & 12 & 14 & 09 & 16 & 08 & 06 & 14 & 15 & 25 & 21 & 20 & 14 & 13 & 10 \\
\end{array}
\]

- **Gaussian**
- **Weibull**

\[ H_1 : R_i = n + m_i \]
**Decision-Theoretic Approach: Gaussian Distribution**

- For Gaussian distribution, Likelihood ratio test will be:

\[
\Lambda(R) = \prod_{i=1}^{N} \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(R_i - m_i)^2}{2\sigma^2} \right)
\]

Equation # 1

\[
\ln(\Lambda(R)) = \sum_{i=1}^{N} \left( \frac{R_i^2 - n^2}{2\sigma^2} \right)
\]

Equation # 2

- Final equation for the Likelihood Ratio test is and threshold on the left is calculated through Neyman-Pearson method

\[
\ln \frac{\eta_{1}}{\eta_{0}} \overset{\sim}{\sim} \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\{ R_i^2 - n^2 \right\}
\]

Equation # 3
For Weibull distribution Likelihood ratio test will be:

\[
\Lambda(R_i) = \prod_{i=1}^{N} \frac{\left(\frac{\gamma}{\alpha}\right)\left(\frac{R_i - m_i}{\alpha}\right)^{(\gamma-1)}e^{-\left(\frac{R_i - m_i}{\alpha}\right)\gamma}}{\left(\frac{\gamma}{\alpha}\right)\left(\frac{R_i}{\alpha}\right)^{(\gamma-1)}e^{-\left(\frac{R_i}{\alpha}\right)\gamma}}}
\]

Final equation for the Likelihood Ratio test is:

\[
\ln\left(\Lambda(R_i)\right) = \sum_{i=1}^{N} \left\{ \left(\gamma - 1\right) \ln\left(\frac{n}{R_i}\right) + \left(\frac{R_i - n}{\alpha}\right)^{\gamma} \right\}
\]
if $\eta_1 < \eta$

Observation $\omega$ is anomalous, add $\omega$'s timestamp to the array of events $\psi$;

elseif $\eta < \eta_0$ then

Observation $\omega$ is not anomalous;

Update the training dataset with $\omega$, discard the oldest entry, recalculate $\sigma_{tr}$ and $\eta_1$;

else Not enough information to make a decision;

Increment $\tau_s$ and $\tau_e$;

end

Analyze $\psi$ and combine consecutive anomalous windows defining unique events;
Summary of Contribution

- Dataset Selection
- Dataset Labeling
- New Event Detection Algorithm
- Comparative Accuracy Evaluation with Existing Techniques

Conclusion and Summary of Contribution

References
Comparative Evaluation: ROC Results

- Detection accuracy is compared in terms of ROC curves
- We draw ROC curves with *True Positive rate* on Y-axis and *False-Alarm rate* on X-axis
- Our proposed technique outperforms the competing anomalous event detection techniques

  - We observe Weibull distribution on DESY & NSLABS paths, so we applied Weibull based DTA model on them.
  - Similarly we observe Gaussian distribution on UTORONTO & FZK paths, so we applied Gaussian based DTA model on them.
ROC Results: SLAC to (FZK & SDSC)
ROC Results: SLAC to (NSLABS & DESY)

(d) NSLABS

(e) DESY
### Comparative Evaluation: Delay Detection Results

* Delay in terms of number of measurements required to raise an alarm

* Number of events detected (INF) Infinity

<table>
<thead>
<tr>
<th>Path</th>
<th>DTA</th>
<th>PLA</th>
<th>AFD</th>
<th>HWF</th>
<th>KAL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>UTORONTO</td>
<td>35</td>
<td>9.77</td>
<td>23</td>
<td>4.98</td>
<td>4</td>
<td>53.25</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>310</td>
<td>4</td>
<td>4.75</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
<td>7.25</td>
<td>1</td>
<td>2.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>280</td>
<td>0</td>
<td>INF</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>DESY</td>
<td>30</td>
<td>5.77</td>
<td>14</td>
<td>12.60</td>
<td>1</td>
<td>47.43</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>INF</td>
<td>0</td>
<td>INF</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>SDSC</td>
<td>5</td>
<td>7.60</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>212</td>
<td>0</td>
<td>INF</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>FZK</td>
<td>17</td>
<td>22.13</td>
<td>7</td>
<td>28.53</td>
<td>3</td>
<td>110.53</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>664</td>
<td>0</td>
<td>INF</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>TRIUMF</td>
<td>2</td>
<td>3.7</td>
<td>3</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>20</td>
<td>0</td>
<td>INF</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NSLAB</td>
<td>4</td>
<td>0.00</td>
<td>2</td>
<td>2.33</td>
<td>2</td>
<td>45.66</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>47</td>
<td>0</td>
<td>INF</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>SWITCH</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>47</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
### Comparative Evaluation: Delay Detection Results

**$\varepsilon$** Delay in terms of number of measurements required to raise an alarm

**#** Number of events detected **(INF) Infinity**

<table>
<thead>
<tr>
<th>Path</th>
<th>DTA</th>
<th>PLA</th>
<th>AFD</th>
<th>HWF</th>
<th>KAL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>$\varepsilon$</td>
<td>#</td>
<td>$\varepsilon$</td>
<td>#</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>UTORONTO</td>
<td>35</td>
<td>9.77</td>
<td>23</td>
<td>4.98</td>
<td>4</td>
<td>53.25</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
<td>7.25</td>
<td>1</td>
<td>2.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>DESY</td>
<td>30</td>
<td>5.77</td>
<td>14</td>
<td>12.60</td>
<td>1</td>
<td>47.43</td>
</tr>
<tr>
<td>SDSC</td>
<td>5</td>
<td>7.60</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>FZK</td>
<td>17</td>
<td>22.13</td>
<td>7</td>
<td>28.53</td>
<td>3</td>
<td>110.53</td>
</tr>
<tr>
<td>TRIUMF</td>
<td>2</td>
<td>3.7</td>
<td>3</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>NSLAB</td>
<td>4</td>
<td>0.00</td>
<td>2</td>
<td>2.33</td>
<td>2</td>
<td>45.66</td>
</tr>
<tr>
<td>SWITCH</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
</tbody>
</table>
### Comparative Evaluation: Delay Detection Results

Delay in terms of number of measurements required to raise an alarm

- # Number of events detected
- INF Infinity

<table>
<thead>
<tr>
<th>Path</th>
<th>DTA</th>
<th>PLA</th>
<th>AFD</th>
<th>HWF</th>
<th>KAL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>$\mathcal{E}$</td>
<td>#</td>
<td>$\mathcal{E}$</td>
<td>#</td>
<td>$\mathcal{E}$</td>
</tr>
<tr>
<td>UTORONTO</td>
<td>35</td>
<td>9.77</td>
<td>23</td>
<td>4.98</td>
<td>4</td>
<td>53.25</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
<td>7.25</td>
<td>1</td>
<td>2.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>DESY</td>
<td>30</td>
<td>5.77</td>
<td>14</td>
<td>12.60</td>
<td>1</td>
<td>47.43</td>
</tr>
<tr>
<td>SDSC</td>
<td>5</td>
<td>7.60</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>FZK</td>
<td>17</td>
<td>22.13</td>
<td>7</td>
<td>28.53</td>
<td>3</td>
<td>110.53</td>
</tr>
<tr>
<td>TRIUMF</td>
<td>2</td>
<td>3.7</td>
<td>3</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>NSLAB</td>
<td>4</td>
<td>0.00</td>
<td>2</td>
<td>2.33</td>
<td>2</td>
<td>45.66</td>
</tr>
<tr>
<td>SWITCH</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
</tbody>
</table>
## Comparative Evaluation: Delay Detection Results

\[ \mathcal{E} \quad \text{Delay in terms of number of measurements required to raise an alarm} \]

\# \quad \text{Number of events detected} \quad (\text{INF}) \quad \text{Infinity}

<table>
<thead>
<tr>
<th>Path</th>
<th>DTA</th>
<th>PLA</th>
<th>AFD</th>
<th>HWF</th>
<th>KAL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>\mathcal{E}</td>
<td>#</td>
<td>\mathcal{E}</td>
<td>#</td>
<td>\mathcal{E}</td>
</tr>
<tr>
<td>UTORONTO</td>
<td>35</td>
<td>9.77</td>
<td>23</td>
<td>4.98</td>
<td>4</td>
<td>53.25</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
<td>7.25</td>
<td>1</td>
<td>2.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>DESY</td>
<td>30</td>
<td>5.77</td>
<td>14</td>
<td>12.60</td>
<td>1</td>
<td>47.43</td>
</tr>
<tr>
<td>SDSC</td>
<td>5</td>
<td>7.60</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>FZK</td>
<td>17</td>
<td>22.13</td>
<td>7</td>
<td>28.53</td>
<td>3</td>
<td>110.53</td>
</tr>
<tr>
<td>TRIUMF</td>
<td>2</td>
<td>3.7</td>
<td>3</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>NSLAB</td>
<td>4</td>
<td>0.00</td>
<td>2</td>
<td>2.33</td>
<td>2</td>
<td>45.66</td>
</tr>
<tr>
<td>SWITCH</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
</tbody>
</table>
Comparative Evaluation: Delay Detection Results

\( \mathcal{E} \)  Delay in terms of number of measurements required to raise an alarm

\#  Number of events detected  (INF) Infinity

<table>
<thead>
<tr>
<th>Path</th>
<th>DTA</th>
<th>PLA</th>
<th>AFD</th>
<th>HWF</th>
<th>KAL</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>( \mathcal{E} )</td>
<td>#</td>
<td>( \mathcal{E} )</td>
<td>#</td>
<td>( \mathcal{E} )</td>
</tr>
<tr>
<td>UTORONTO</td>
<td>35</td>
<td>9.77</td>
<td>23</td>
<td>4.98</td>
<td>4</td>
<td>53.25</td>
</tr>
<tr>
<td>CERN</td>
<td>8</td>
<td>7.25</td>
<td>1</td>
<td>2.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>DESY</td>
<td>30</td>
<td>5.77</td>
<td>14</td>
<td>12.60</td>
<td>1</td>
<td>47.43</td>
</tr>
<tr>
<td>SDSC</td>
<td>5</td>
<td>7.60</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>FZK</td>
<td>17</td>
<td>22.13</td>
<td>7</td>
<td>28.53</td>
<td>3</td>
<td>110.53</td>
</tr>
<tr>
<td>TRIUMF</td>
<td>2</td>
<td>3.7</td>
<td>3</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>NSLAB</td>
<td>4</td>
<td>0.00</td>
<td>2</td>
<td>2.33</td>
<td>2</td>
<td>45.66</td>
</tr>
<tr>
<td>SWITCH</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>2.33</td>
<td>0</td>
<td>INF</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Motivation
- Existing Work
  - Contemporary Anomalous Event Detectors
    - Most existing algorithms cannot be adopted
    - Algorithms adopted for Path Event Detection
- Research Contributions
  - Dataset Selection
  - Dataset Labeling
  - New Event Detection Algorithm
  - Comparative Accuracy Evaluation with Existing Techniques

**Conclusion and Summary of Contribution**

- References
**PathChirp** is an appropriate tool for path event detection and *available bandwidth* is a suitable metric

- *TCP Reno* makes Iperf & Thrulay un-suitable for high speed links

**Appropriate definition of event is:**

- Magnitude of change = 0.5μ to 1.5μ
- Duration of Change = 3 Hours

**Irrespective of the bandwidth estimates’ distribution, a Decision-Theoretic event detector outperforms competing algorithms in terms of Detection Accuracy and Detection Delays**
References


References


15. L. Cottrell et al. “Evaluation of techniques to detect significant network performance problems using end-to-end active network measurements,”

## Magnitude of Change: Intuition

### N = 7

<table>
<thead>
<tr>
<th>200</th>
<th>211</th>
<th>220</th>
<th>210</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>200</th>
<th>205</th>
<th>216</th>
<th>210</th>
<th>220</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>200</td>
<td>210</td>
<td>211</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### N = 5

<table>
<thead>
<tr>
<th>200</th>
<th>211</th>
<th>220</th>
<th>210</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>200</th>
<th>205</th>
<th>216</th>
<th>210</th>
<th>220</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>210</td>
<td>220</td>
<td>210</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Why do we need this?
### Example of median filter, $n = 3$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
### Example of median filter, $n = 3$

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>20</th>
<th>4</th>
<th>5</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- Example of median filter, $n = 3$

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>20</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>20</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- Example of median filter, n = 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>20</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>20</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example of median filter, $n = 3$

\[
\begin{array}{cccc}
4 & 3 & 20 & 4 & 5 & 3 \\
3 & 4 & 20 & & & \\
3 & 4 & 20 & & & \\
4 & 5 & 20 & & & \\
3 & 4 & 5 & & & \\
3 & 5 & & & & \\
& & & 3 & 5 & \\
\end{array}
\]

\[n \geq 2\delta \nu\]

- $\delta$ is the minimum event length
- $\nu$ is the average number of measurements per hour
We make subsets of dataset by decimating values away from the mean such that:

\[ \left( \gamma * E\left( M\left( X_n \right) \right), \kappa * E\left( M\left( X_n \right) \right) \right) \]
Magnitude of Change: Markov Chain Analysis

B = Benign
A = Anomalous


---
Magnitude of Change: Markov Chain Analysis

B = Benign
A = Anomalous
Magnitude of Change: Markov Chain Analysis

B = Benign
A = Anomalous
Magnitude of Change: Markov Chain Analysis

B = Benign
A = Anomalous
Magnitude of Change: Markov Chain Analysis

B = Benign
A = Anomalous

Benign Dataset
We make subsets of dataset by dissipating values away from the mean such that:

\[
\left( \gamma \ast \mathbb{E}\left( M\left( X_n \right) \right), \kappa \ast \mathbb{E}\left( M\left( X_n \right) \right) \right)
\]
compute $n = 2\delta v$;

Apply $n$–tap median filter to $\Omega$ to obtain $\Omega_f$ and consequently $\mu_f$

for $\{w_i \in \Omega | 1 \leq i \leq N\}$ do

Compute $\mu_\Delta$

if $\left(\frac{\mu_\Delta}{\mu_f} \leq \text{left} \leq \text{right}\right)$ then

Mark as normal observation;

else

Mark as ananomalous window and add to $\tau$;

end

for all alerts in $\tau$; do

if required coalesce alert–windows considering $\delta$ to identify unique observations with adjusted boundaries;

end
Decision-Theoretic Approach: Detection Algorithm

**Data:** Array of performance measurements $\Omega$, False positive rate $\alpha$, Detection rate $\beta$, window size $\rho$, Initial duration for training dataset $\delta$ and width of median filter $\nu$.

**Result:** Array of timestamp-brackets $\psi$ classifying windows as containing events.

Apply low–pass median filter of width $\nu$ to obtain $\Omega_{tr}$;

Compute $\mu_{tr}$ and $\sigma_{tr}$ for \( \{ \omega_t \in \Omega_{tr} \mid t_0 < t < t_0 + \delta \} \);

Let threshold \( \eta_1 = \frac{\beta \sigma^2}{\alpha} \), \( \eta_0 = \frac{1 - \beta}{1 - \alpha} \) and \( t_0 = 0 \);

/*determine the baseline*/
/*determine the baseline*/

Let $\tau_s = t_0 + \delta - \nu$ and $\tau_e = t_0 + \delta$;

for \( \{ \omega \in \Omega \} \) do

Let $x_1 = \text{rand}()$, $x_2 = \text{rand}()$ and $n = \sqrt{-2 \ln(x_1)} \cdot \sin(2\pi x_2) \cdot \sigma$;

\[ R = \text{median}\{ \omega_i \mid \tau_s \leq i \leq \tau_e \} ; \]

Compute $\eta = \frac{R^2 - n^2}{2\sigma_{tr}^2}$;
Algorithms adopted for Path Event Detection

- **Plateau Algorithm (PL)** by Logg et al.; Stanford Linear Accelerator Center (SLAC) [1]
  - Uses two buffers:
    - History Buffer
    - Trigger Buffer
  - Calculates the mean & standard deviation of both buffers
  - Any value away from mean by 2 standard deviations is considered as anomalous

- **Adaptive Fault Detector (AFD)** by Hajji; IBM Business Consulting [3]
  - Models traffic as a K-Variate Gaussian Random Variable
  - Uses an increment process to observe difference between consecutives values
  - Detection is performed in two phases
    - Training a baseline model
    - Flagging sudden changes using the Likelihood Ratio Test
Plateau Algorithm

Gaussian Probability Distribution Function

Anomalous Region

Confidence Band

Anomalous Region
Adaptive Fault Detector (AFD) Algorithm

Assumes Multiple (K-variate) Gaussian Distributions

Anomalous
Or
Benign

\[ Q(\theta | \theta') = \sum_{i=1}^{N} \sum_{k=1}^{K} w_{ki} \log(f_{ik}(\theta)) \]

\[ w_{ki} = \frac{\pi'_k f_{ik}(\theta')}{\sum_{k=1}^{K} \pi'_k f_{ik}(\theta')} \]

\[ f_{ik} = \frac{1}{\sqrt{2\pi\sigma_k}} \exp \left( -\frac{(x_i - m_k)^2}{2\sigma_k^2} \right). \]
Algorithms adopted for Path Event Detection

- **Holt-Winter Forecasting Algorithm (HWF)** by Jake T Brutlag.; USENIX [4]
  - detects aberrant behavior in time series for network monitoring using:
    - A baseline behavior
    - A linear trend
    - Seasonal effect
  - Uses exponential smoothing to update variables
  - Measures confidence bands
  - If a value lies outside this interval, it is flagged as an event

- **Kalman Filter (KF) Detector** by Augustin et al.; Intel Research [3]
  - Kalman Filter is used to extract path behavior
  - Difference between the expected behavior & actual behavior is called residual
  - Residual is:
    - Compared with User Defined Thresholds
    - Compared with Local Variance & Global Variance
    - Used in a Likelihood Ratio
  - An event is flagged if one of the above conditions is true
Holt Winter Forecasting Algorithm

12, 13, 14, 15, 16, 17

Baseline
12,13,14,15,16,17

Linear Trend
+1

Seasonal Trend
+185, +285

\[(y'_t - \delta_+ \cdot d_{t-m}, y'_t + \delta_- \cdot d_{t-m})\]

Calculates the confidence bands
Limitations of Existing Algorithms

- For accurate anomaly detection, an anomaly detectors should consider the inherent statistical nature of the data

- But:
  - **Plateau Algorithm**
    - Does not consider random spikes in data
  - **Adaptive Fault Detection**
    - Assumes multiple Gaussians
    - Assumes Gaussian noise
    - Injects the means to his model to flag an alarm, result into large detection delays
  - **Holt-Winter Forecasting**
    - Does not consider statistical distribution of data
    - Does not filter random spikes, as a result take noise a seasonal trend
  - **Kalman Filter**
    - Assumes pure Gaussian distribution, does not cater for random fluctuations
    - Assumes additive Gaussian noise
    - Does not provide a method to initiate model parameters

- Based on the above limitations, we anticipated that these algorithms will not be accurate in terms of true positive and false alarm rates

- Our results discussed later will substantiate this claim
The filtered data is then evaluated to quantify the mean and variance of the baseline behavior

\[ |x - \mu| = s \]

We divide the \( s \) in 4 quarters using difference from the mean
The filtered data is then evaluated to quantify the mean and variance of the baseline behavior

\[ |x - \mu| = s \]

We divide the \( s \) in 4 quarters using difference from the mean

| Percentage of difference of Available Bandwidth w.r.t. Mean (Slac_Utoronto) |
|-----------------------------|-----------------------------|
| 4.8752862                   | 1.1227932                   |
| 26.397951                   | 67.603969                   |

| Percentage of difference of Available Bandwidth w.r.t. Mean (Slac_Desy) |
|-----------------------------|-----------------------------|
| 5.2285555                  | 0.3752403                   |
| 12.745767                  | 81.650437                   |

Legend:
- \( \text{Diff} < 25\% \)
- \( 25\% < \text{Diff} \leq 50\% \)
- \( 50\% < \text{Diff} \leq 75\% \)
- \( 75\% < \text{Diff} \leq 100\% \)
Magnitude of Change: Mean as a Qualitative Measure

- The filtered data is then evaluated to quantify the mean and variance of the baseline behavior
- \(|x - \mu| = s\)
- We divide the \(s\) in 4 quarters using difference from the mean

6% values with a difference of 0.5 \(\mu\)