AnomVis is a comprehensive anomaly detection and data visualization tool that streamlines human operators’ ability to identify anomalies in networks and make efficient, informed decisions.

## Abstract

AnomVis is a comprehensive anomaly detection and data visualization tool that streamlines human operators’ ability to identify anomalies in networks and make efficient, informed decisions.

## Motivation

**Large amounts of data**
- Comcast has 1-2 GB of internal network data per day per data center

**Need to monitor system behavior**
- Server crashes disrupt company activities
- Attacks can cost companies millions

**Lack of user focused applications**
- Time consuming and difficult to identify issues
- Busy visualizations can overwhelm users

## Our Application

Parses, processes, and analyzes data in parallel
- Classifies data points and records trends
Depicts results using cognitive metaphor theory
- Focuses on color, size, and refined views

## Architecture

- Network Logger
- S3 Bucket

**Data Processing Server** running a daily job on EC2
- Parse relevant metrics
- Format data for processing
- Route for metric / endpoint
- D3 data transformations
- Analyze with K-Means algorithm

**Web Application** hosted on Elastic Beanstalk
- User Interaction

## User Interface

**Month View**
- Sidebar calendar
- Days colored by anomalousness

**Day View**
- Slider to select hours
- Details about selected point in the graph

## Data Processing

- Apache Spark distributes computation and clustering over available cores
- Backend uploads Spark processing results to S3
- Maintains historical trend information
- Formats data for the web application to display

## Evaluation

**A/B Testing Results:**
- 20% more of B determined the anomalous hour
- 50% faster average than A on average

## Future Work

- Setup daily analysis job
- Read directly from log files
- Expand infrastructure to handle real-time resource load
- Further analysis of data