The Shape of View
A Real-Time Alerting Framework for Video Viewership Anomalies

Antonis Manousis, Harshil Shah*, Henri Milner*,
Yan Li*, Hui Zhang*, Vyas Sekar

Carnegie Mellon University, *Conviva Inc
Video streaming services critically rely on telemetry

Effective ad-placement in stream can reduce content abandonment by half

Avoiding 1 buffering event can increase watch time by 63%

Content providers deploy workflows to monitor engagement and QoE
Yet, existing monitoring workflows still have blind spots

Our team has resolved streaming issues with #HBOMax. You may need to close and reopen HBO Max, or sign out and sign back in.

Changes in viewership patterns can be indicators of anomalies!
Our Goal:

Real-time alerting on video viewership anomalies
Proteas: Real-time alerting for viewership anomalies

**Viewership**: Timeseries of the number of viewers.

**Viewership Anomaly**: Unexpected change in viewership that requires technical intervention

- **Detection**
  - NYC Viewers on Comcast
  - SF Viewers on Apple TV
  - Chicago Viewers on Roku TV
  - Viewers on Roku TV

- **Subpopulation of viewers**

- **Diagnosis**
  - Viewers on Roku TV
  - Potential root causes:
    - Device outage
    - Buggy Firmware

- **Summarize detected anomalies**

- **Identify candidate root causes**

- **Anomaly detection across subpopulations of viewers**
Challenge 1: Contextual viewership anomalies

Context is group-specific and can only be captured with model-based techniques.
Challenge 2: Viewership is non-stationary

Model-based anomaly detection techniques commonly assume stationarity

Statistical properties of viewership change over time

Large temporal variability of viewership indicates non-stationarity

2x more viewers

Viewership

00:00 04:00 08:00 12:00 16:00 20:00

Minute of Day

Sunday #1
Sunday #2
Sunday #3
Sunday #4
Sunday #5
Sunday #6
Insight: Shape of daily viewership remains consistent over time

Remove Magnitude Variability
Insight: Shape of daily viewership remains consistent over time

**Approach:** Use *shape* of viewership as baseline for anomaly detection

Remove Magnitude Variability
Modeling requirements for shape of viewership

1. Generality to the functions it can model
   o Viewership function is non-linear, non-continuous function

2. Robust to Outliers
   o Ideally provides confidence bounds around prediction

3. Fast Predictions without need for maintaining much history
   o Detection happens in parallel across 1000s of aggregations

4. Automated Modeling
   o Minimize human intervention, configuration cycles
### Candidate modeling approaches

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Automated</th>
<th>Robust</th>
<th>Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decomposition w/ basis functions</td>
<td>❌</td>
<td>✔️</td>
<td>❌</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Model the shape of viewership as a combination of polynomials, wavelets, functions of the Fourier base, etc.

![Wavelet examples](wavelet_examples.png)
### Candidate modeling approaches

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Automated</th>
<th>Robust</th>
<th>Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decomposition w/ basis functions</td>
<td>❌</td>
<td>✔️</td>
<td>❌</td>
<td>✔️</td>
</tr>
<tr>
<td>Point-wise estimations</td>
<td>✔️</td>
<td>✔️</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>

- **Rely on summary statistics of viewership at different timestamps**

- **Maintain summary statistics of viewership at each timestamp**
Candidate modeling approaches

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Automated</th>
<th>Robust</th>
<th>Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decomposition w/ basis functions</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Point-wise estimations</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gaussian processes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Model shape of viewership as a multivariate probability distribution

Viewership at time $t$ is sampled from a random variable

Observations are correlated

$$\Sigma_{i,j}$$
Candidate modeling approaches

<table>
<thead>
<tr>
<th>Framework</th>
<th>General</th>
<th>Automated</th>
<th>Robust</th>
<th>Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decomposition w/ basis functions</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Point-wise estimations</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Gaussian processes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Model shape of viewership as a multivariate probability distribution
Proteas: Real-time alerting for viewership anomalies

**Viewership**: Timeseries of the number of viewers.

**Viewership Anomaly**: Unexpected change in viewership that requires technical intervention

---

**Detection**
- NYC Viewers on Comcast
- SF Viewers on Apple TV
- Chicago Viewers on Roku TV
- Viewers on Roku TV

**Diagnosis**
- Summarize detected anomalies
- Identify candidate root causes
- Potential root causes
  - Device outage
  - Buggy Firmware

**Anomaly detection across subpopulations of viewers**
One logical incident produces redundant anomalies!

Viewership groups are not disjoint sets! Same anomaly manifests in >1 groups!

Viewership groups of 1 feature
- COX
- Verizon
- Comcast
- Apple TV
- Amazon Fire

Viewership groups of 2 features
- COX
- Apple TV
- Verizon
- Apple TV
- Comcast
- Apple TV
- Verizon
- Fire
- Comcast
- Fire

Can we “summarize” the detected anomalies in one root-cause group?
Insight: Anomalies propagate predictably across groups

For every anomalous group, its explanatory power $EP_g$ is the produce of 2 group-specific metrics:

$$EP_g = M^1_g \times M^2_g$$

$$M^1_g = \frac{\#\text{anomDescendants}_g}{\#\text{allDescendants}_g}$$

Higher ratio of anomalous descendants $\rightarrow$ better $EP_g$
Insight: Anomalies propagate predictably across groups

For every anomalous group, its explanatory power $EP_g$ is the produce of 2 group-specific metrics:

$$EP_g = M_g^1 * M_g^2$$

$$M_g^1 = \frac{\#\text{anomDescendants}_g}{\#\text{allDescendants}_g}$$  

Higher ratio of anomalous descendants $\Rightarrow$ better $EP_g$

$$M_g^2 = \frac{|\hat{V}_g - V_g^{obs}|}{\max\{|\hat{V}_{g'} - V_{g'}^{obs}|\}}$$

Highest deviation from predicted viewership $\Rightarrow$ better $EP_g$

The top-k groups ordered by EP are the best candidates for explaining the anomalies
Identifying root causes using anomaly signatures

Anomaly Signature

Spatial Dimension
Patterns in hierarchical ordering of anomalous groups

Temporal Dimension
1. Start time
2. Duration
3. Shape

Anomalies that don’t match with any signature in the library are sent to an analyst for further analysis.
Evaluation methodology

**Dataset:** 13 weeks of viewership from 3 large content providers

**2 Baselines:** Twitter’s anomaly detection, Surus (Netflix)

**Ground Truth:** Original dataset unlabeled and ground truth was lacking.

- **Twitter-based sentiment analysis**
  - Sentiment analysis on providers’ help-desk Twitter profiles

- **Public downtime records**
  - Service uptime reports from public databases

- **Holiday calendars**
  - Singular Events and Holidays

- **Video QoE alerts**
  - Alerts from existing video QoE workflows

- **Blind review by experts**
  - Run survey with 4 video analysts asking to mark anomalous incidents
Accurate (>86% True Positive Rate)
No Detected False Negatives
Few False Positives (Precision >0.85)
PROTEAS’ precision is better than that of baselines
Other questions in the paper

1. Does PROTEAS add value to content providers’ existing monitoring workflows?

2. Do analysts trust the alerts issued by PROTEAS?

3. Does summarization result in PROTEAS issuing a manageable number of alerts?

4. What do true positives look like?
   - Breakdown by session attributes
   - Breakdown by source of ground truth
   - Breakdown by anomaly signature
Takeaways

• Real-time alerting of viewership anomalies can complement existing video streaming telemetry workflows

• PROTEAS leverages persistent structural patterns of the shape of viewership to enable accurate and efficient anomaly detection

• PROTEAS efficiently summarizes redundant anomalies and uses anomaly signatures to output manageable number of insightful alerts to content providers.

PROTEAS’ key ideas are being deployed to production

Contact: Antonis Manousis, antonis@cmu.edu