Network Syslog Analysis with DeCorus
Bruno Wassermann, David Ohana, Elliot K. Kolodner, Michal Malka, Eran Raichstein, Ronen Schaffer, Robert Shahla, Moshik Hershcovitch
IBM Research Haifa

**Motivation**
Fast detection of failures in the network infrastructure of a modern cloud service provider.

**Scale**
~10,000 network-devices in a single DC. Reliability team supports up to 100 DCs.

**Data**
Network device syslogs: Up to 1 billion of text-based syslog events in various formats per day.

**Current State**
Matching each log event to a DB of handcrafted regex-based rules. Does not cover new types of errors; produces too many alerts; requires a lot of maintenance.

---

**DeCorus: Detection and Correlation of Unusual Signals**
DeCorus is a general purpose streaming pipeline for online, hierarchical, unsupervised anomaly-detection and root-cause analysis. We apply it on network-device syslogs.

1. **Log-Template Mining**
   - Header Extraction: time, host, severity, event-code, etc..
   - Log templates are extracted from event textual content using DRAIN3 clustering algorithm.
   - Output: Thousands of unique log templates.

2. **Conversion to Time Series**
   - Counting occurrences of each template per network device per 5-minute time interval.
   - Output: Up to 100K of unique time-series (signals).

3. **Univariate Anomaly Detection**
   - Detecting anomalies in each time series independently.
   - Output: Anomaly score per signal per time-interval.

4. **Multivariate Anomaly Detection**
   - Home-grown, topology-based algorithm for scoring anomaly level of the data centre.
   - Output: Anomaly scores per DC & network device; significant anomaly dimensions.

---

**Alerting, Visualisation, Root-Cause Analysis**