GRETA: Graph-based Real-time Event Trend Aggregation

WPI · Portland State University · IBM Research

Olga Poppe, Chuan Lei, Elke A. Rundensteiner, and David Maier

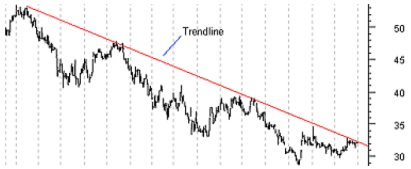
Supported by NSF grants IIS 1018443, IIS 1343620, IIS 1560229, CRI 1305258.

Event Trends in Streaming Applications

Goal: Reliable actionable insights about the stream

Solution: Each event is considered in the context of other events in the stream

Algorithmic Trading



Health Care

Irregular heat rate

Traffic Control

Aggressive driving

Financial Fraud

Circular check kite

Cluster Monitoring

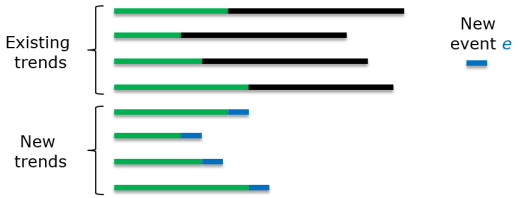
Unbalanced load distribution

Single event = *Single* stock value may be an outlier

Event sequence = Stock trend of *fixed* length may be a local fluctuation

Event trend = Stock trend of *arbitrary* length provides reliable insights

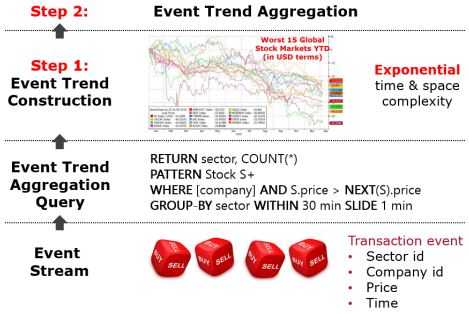
Challenges



- **Exponential number** of trends
- **Arbitrary length** of a trend
- **Complex event inter-dependencies** in a trend

➡ **Exponential** time complexity

Existing Two-Step Approaches



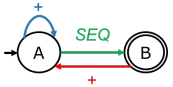
Problem Statement: Real-time response despite exponential costs

Graph-Based Event Trend Aggregation

Nested Kleene Pattern

$$P = (SEQ(A+, B))+$$

Graph Template

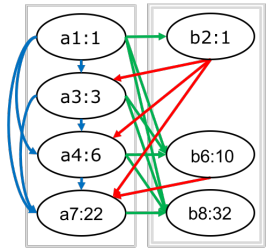


Template captures instructions for runtime graph construction

States are event types

Transitions are event operators

GRETA Graph



Final count: 43

Graph compactly captures all trends

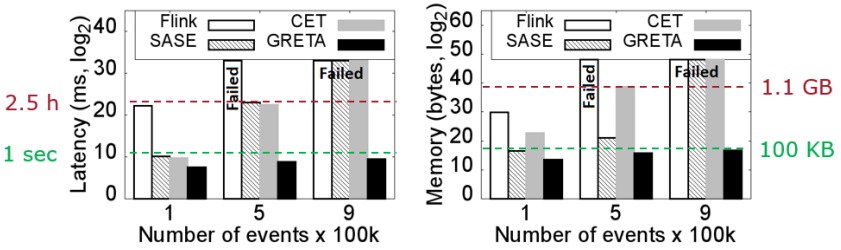
Nodes are matched events
For example, (b2:1) is an event of type B with time stamp 2 and intermediate count 1 (number of trends that end at b2)

Edges connect adjacent events in a trend

Aggregates are incrementally propagated along the edges

Quadratic time & **linear** space complexity

Experiments



GRETA is a win-win solution that achieves **10^4 speed-up** and **10^8 memory reduction** compared to existing approaches

Conclusions

We are the first to compute **aggregation of Kleene closure matches over event streams** with **optimal time complexity**

1. **GRETA graph** compactly encodes all event trends matched by expressive Kleene queries
2. Graph-based event trend aggregation with **quadratic time complexity**