ECRTS 2016
5-8 July
Toulouse, France
Work-in-Progress Session

Inter-Arrival Curves for Multi-Mode and Online Anomaly Detection

Mahmoud Salem, Mark Crowley,
and Sebastian Fischmeister

UNIVERSITY OF WATERLOO

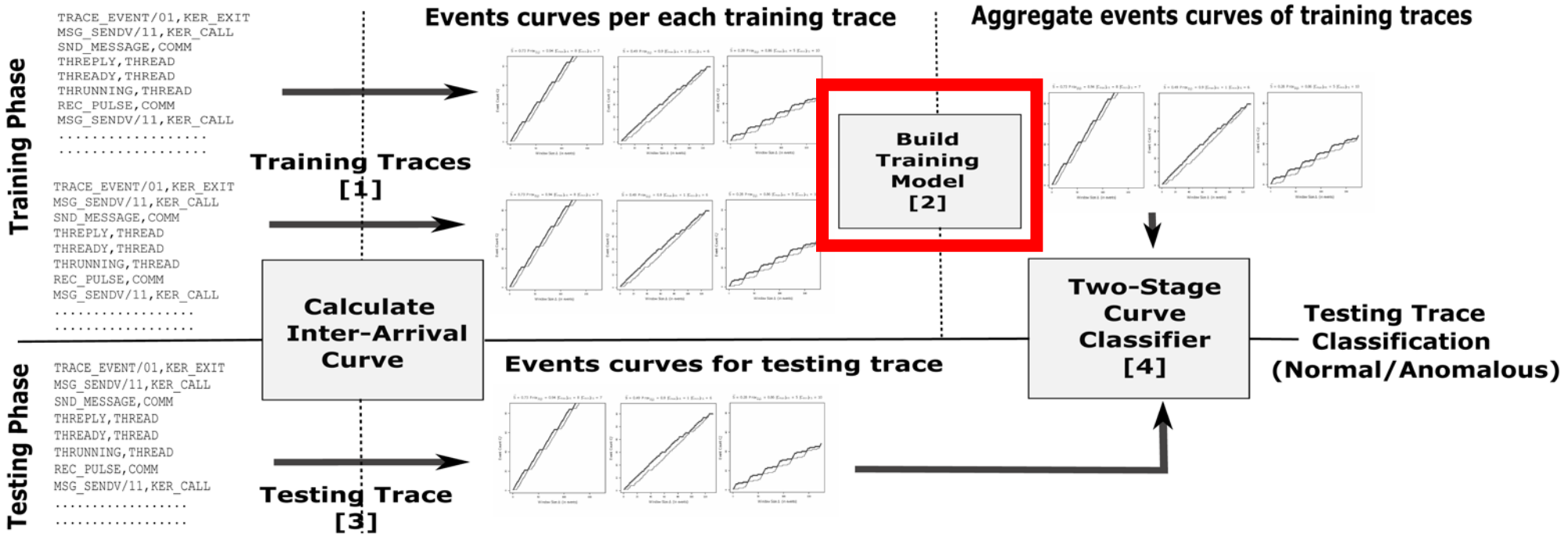
Inter-arrival Curves for Anomaly Detection [1]

- Inter-arrival curves make good features for reasoning about recurrent behavior using event traces
- Promising classification results from an offline anomaly detection framework, however some anomalies go undetected
- Current research interest in online anomaly detection approaches

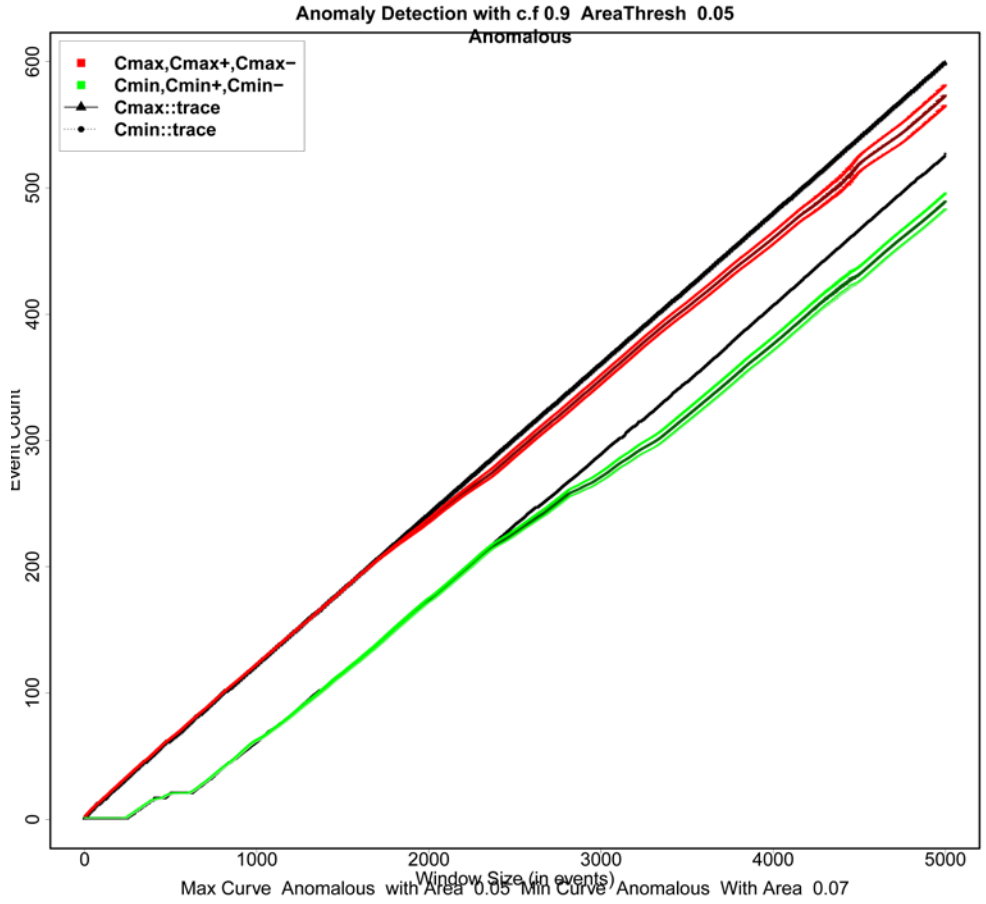
[1] “Anomaly Detection using Inter-Arrival Curves for Real-time Systems” – ECRTS’16

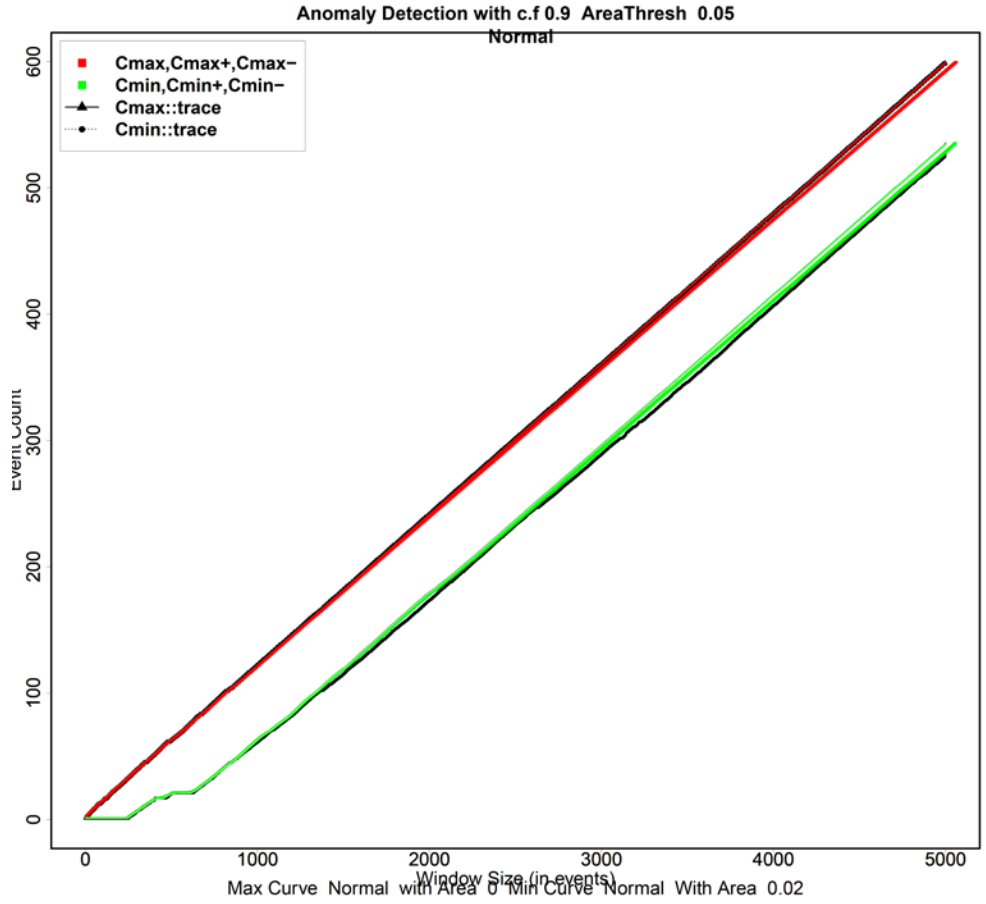
Problem Statement & Approach (1)

"Given a **set of event traces** generated by a **well-specified** system that exhibits **several modes of operations**, check whether a **new trace** from the same system reflects **any of these modes** of operation."



Preliminary Evaluation: Multi-Mode Model

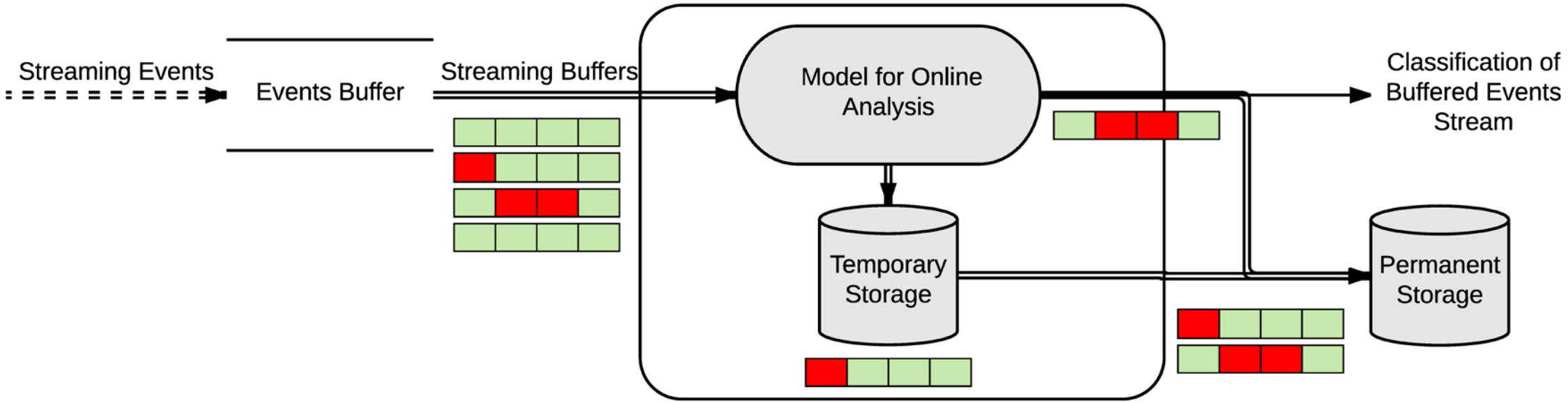




Using a single model falsely flags a normal mode of operation as anomalous.

Problem Statement & Approach (2)

"Given a **set of event traces** generated by a **well-specified system** in a given execution scenario, check **on-the-fly** whether **a stream of events** from the same system originates from the **same execution scenario**."



Preliminary Evaluation: Online Anomaly Detection

- Synthetically stream trace data files
- Using $|T|_{testing} \approx \Delta_{max}$ and $|T|_{testing} << |T|_{training}$

Training Scenario	Normal Testing Scenario	Anomalous Testing Scenario	TPR	FPR
50 files	129 files	185 files	84%	0%

Target Contribution

- Online anomaly detection technique for event traces using inter-arrival curves
- Multi-mode classification framework using inter-arrival curves for improved anomaly detection
- Empirically demonstrate the feasibility and viability of the proposed approaches using event traces from embedded real-time systems

Thank you !

Anomaly Detection Using Inter-Arrival Curves for Real-time Systems

Mahmoud Salem, Mark Crowley, and Sebastian Fischmeister
Department of Electrical and Computer Engineering

UNIVERSITY OF WATERLOO

ABSTRACT

Real-time embedded systems are a significant class of applications, poised to grow even further as automated vehicles and the Internet of Things become a reality. An important problem for these systems is to detect anomalies during operation. Anomaly detection is a form of classification, which can be driven by data collected from the system at execution time.

We propose inter-arrival curves as an analytic modelling technique for discrete event traces. The approach relates to the existing technique of arrival curves. Inter-arrival curves analyze the behavior of events within a trace by providing upper and lower bounds to their inter-arrival occurrence.

MOTIVATION & CHALLENGES

- Why Trace Mining for Real-time Systems ?
  - Safety critical systems must meet requirements and standards, e.g. ISO-26262.
  - Real-time systems generate event traces as part of the normal design.
  - Embedded systems are bespoke and often implement recurrent behavior.
- Challenges Facing Trace Mining for Real-time Systems
  - High computational expense of the analysis of data in event traces.
  - Lack of features specific to real-time systems for effective reasoning.
  - Need for extracted models that allow for easier human interpretation.
  - Accurate classifiers due to the cost of false positives and false negatives.

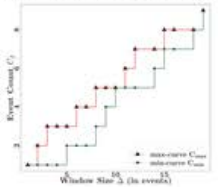
INTER-ARRIVAL CURVES

- A special form of arrival curves that considers the inter-arrival behavior of events.
  - Maps a trace to two types of events ϵ and $\bar{\epsilon}$.
  - Sliding windows positioned to start at ϵ.
  - Cumulative functions in Δ discrete events $f(\Delta)$.

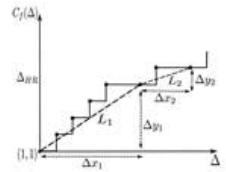
$$\hat{T} = \{ a\ a\ b\ c\ a\ a\ b\ b\ c\ a\ a\ a\ b\ b\ a\ c\ c\ c\ a\ a \}$$

$$T = \{ \bar{\epsilon}\,\bar{\epsilon}\,\bar{\epsilon}\,\epsilon\,\bar{\epsilon}\,\bar{\epsilon}\,\bar{\epsilon}\,\epsilon\,\bar{\epsilon}\,\bar{\epsilon}\,\bar{\epsilon}\,\bar{\epsilon}\,\epsilon\,\bar{\epsilon}\,\bar{\epsilon}\,\bar{\epsilon}\,\epsilon\,\bar{\epsilon} \}$$

Using $T = \hat{T} \uparrow \epsilon$, $C_{min}(T, 4) = 1$ and $C_{max}(T, 4) = 3$



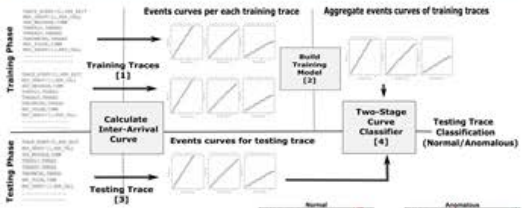
- Curves of Interest
  - Maximum inter-arrival curve C_{max}
  - Minimum inter-arrival curve C_{min}
  - Difference curve $C_{diff} = C_{max} - C_{min}$
- Single-Curve Metrics
  - Steadiness slope calculation.
  - Area under inter-arrival curves.
- Multiple-Curve Metric
  - Proximity of multiple inter-arrival curves.



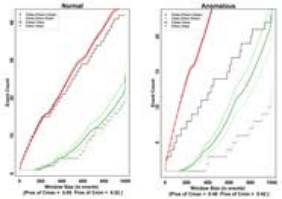
APPLICATIONS

- Offline Anomaly Detection
  - Compute inter-arrival curves using training traces.
  - Build behavioral training model from inter-arrival curves.
  - Post-mortem computation of inter-arrival curves from testing traces.
  - Two-stage classifier to detect and quantify the deviation of a curve from normal model.
- Recurrent Pattern Mining
  - Extract recurrent behavior using inter-arrival curves.
  - Reasoning for recurrent behavior characteristics for anomaly detection purposes.

OFFLINE ANOMALY DETECTION

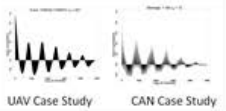


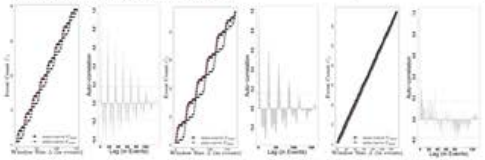
- Non-parametric statistical test inspects whether curves have a similar shape.
- Proximity metric defined as ratio of the area under two curves.
- Apply threshold on proximity metric of a curve from the testing trace against the corresponding curve from the training model.



RECURRENT PATTERN MINING

- Recurrent pattern defined by repeating points of $C_{max} \approx C_{min}$ or $C_{diff} \approx 0$.
- Apply autocorrelation of C_{diff} against lag Δ.
- Statistically significant values separated by approximately equal lag Δ_p indicate recurrence interval Δ_p within a trace T.



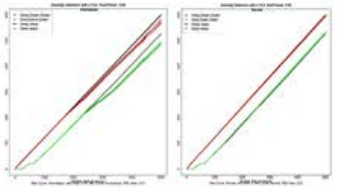


WORK-IN-PROGRESS

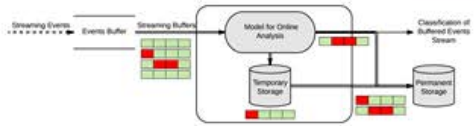
- Multi-Mode Classification
  - The single abstract model fails to detect all anomalies.
  - Propose to model the normal behavior as multiple sub-models.
  - Each sub-model corresponds to one or more modes of operation.
  - Multi-class classification detects the anomalous mode of operation.
- Online Anomaly Detection
  - Detect anomalies in event streams on-the-fly.
  - Overwhelming data generation rates require accurate single-pass methods.

MULTI-MODE CLASSIFICATION

- The device under test mimics data collection system having modes of operation as "data acquisition, data compression, data checksum, etc."
- Having a single-model falsely flags a normal mode of operation as anomalous.
- Multi-mode classification accurately labels modes of operation as normal.



ONLINE ANOMALY DETECTION



- Model trained offline using complete event traces.
- Streaming of events is simulated using complete test traces.
- For efficiency, minimize window size Δ_{max}, yet get acceptable ROC curve.
- False positives unnecessarily consume the limited data storage.

CONCLUSION

- Inter-arrival curves are a special form of well-known arrival curves.
- Inter-arrival curves provide good features for reasoning about recurrent behavior.
- Quantifiable shape-based metrics automate classification for anomaly detection.
- Concepts widely applicable:
  - Various types of discrete event traces, e.g., OS events, function calls, CAN messages.
  - Different computation functions and metrics for reasoning.
  - Various curve aggregation techniques for building robust models.
- Several promising extensions and applications.