Tracing Optimization for Performance Modeling and Regression Detection

Kaveh Shahedi, Heng Li
Polytechnique Montréal

December 2023

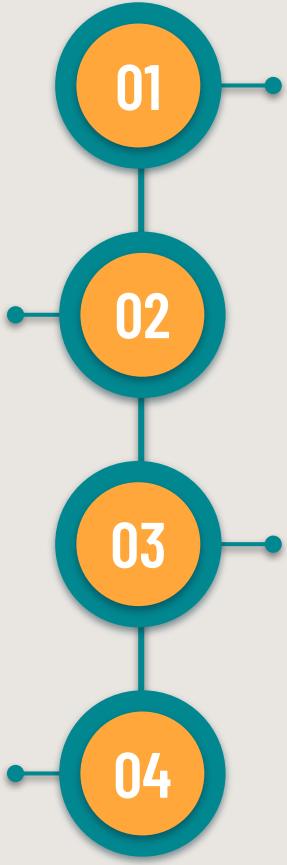
TABLE OF CONTENTS

01

The Problem

The right balance between
accuracy and overhead

Performance Modeling



Observation

Collect various internal/external behaviors (e.g., inputs, functions' executions, etc.)

Modeling

Correlate the behavior characteristics to the program's performance

Prediction

Estimate the performance of the system based on new observations

Evaluation

Assess the performance of the program, and detect any unexpected performance regression

The Trade-Off...

Performance Model's Accuracy

Fully tracing may result in very accurate performance models

- R^2 *Score > 95%*
- *Mean Prediction Error < 5%*

Tracing Overhead

Tracing of high-computational applications is quite expensive

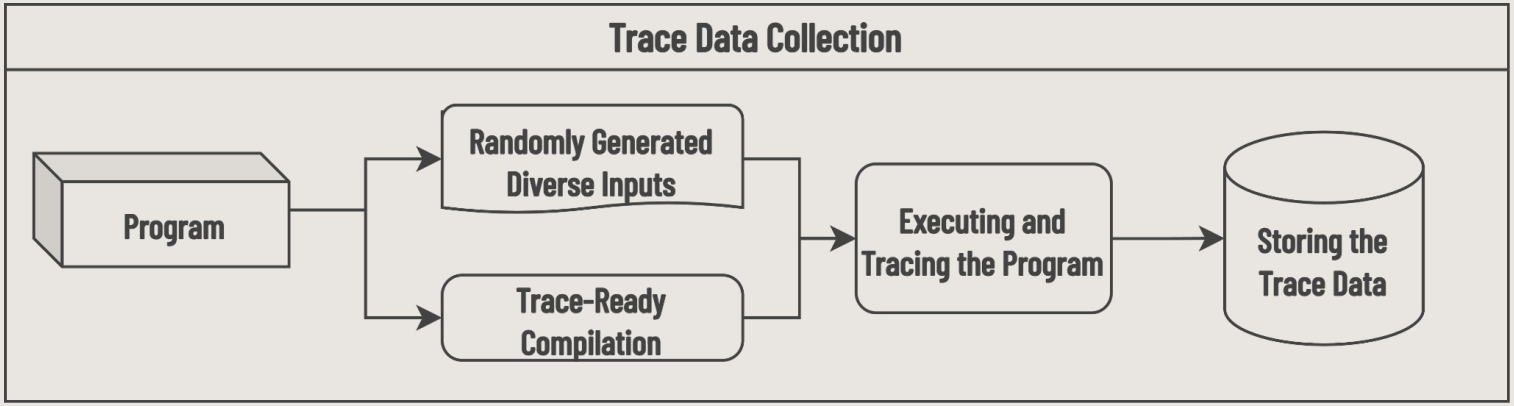
- *Mean Execution Time Overhead >> 50%*
- *Mean Storage Usage Overhead >> 1000%*

02

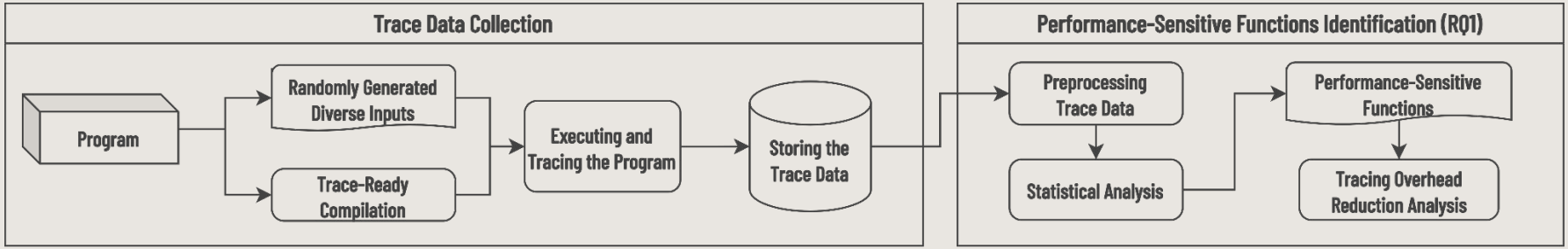
Methodology

Let’s trace only performance-sensitive functions

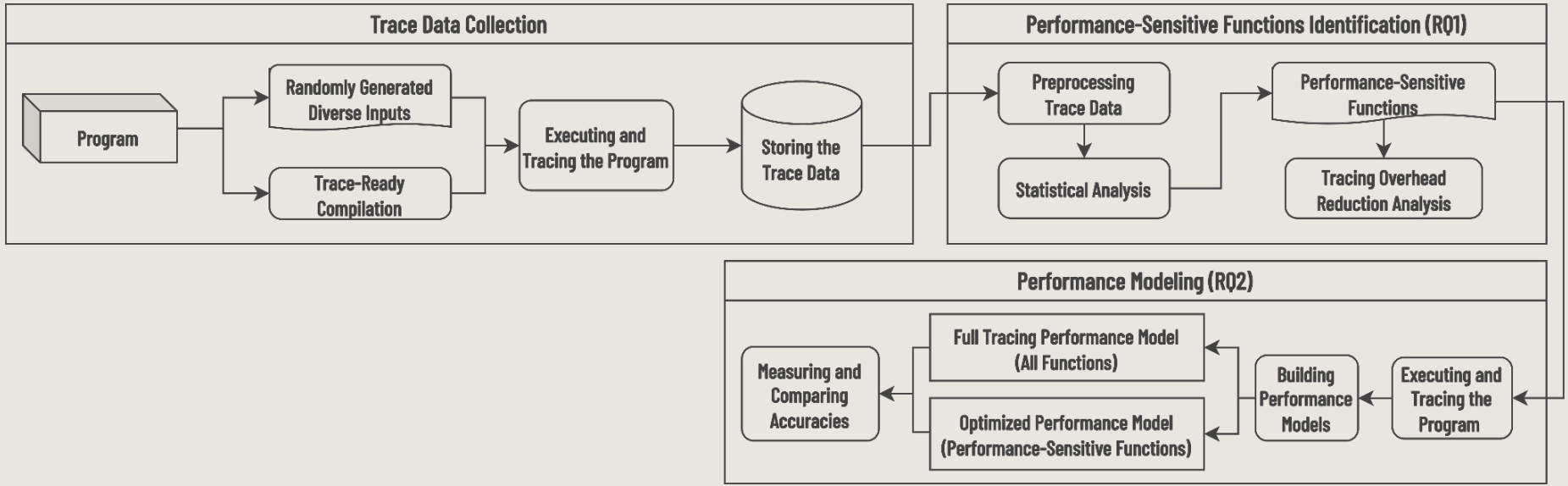
Methodology - Step 1



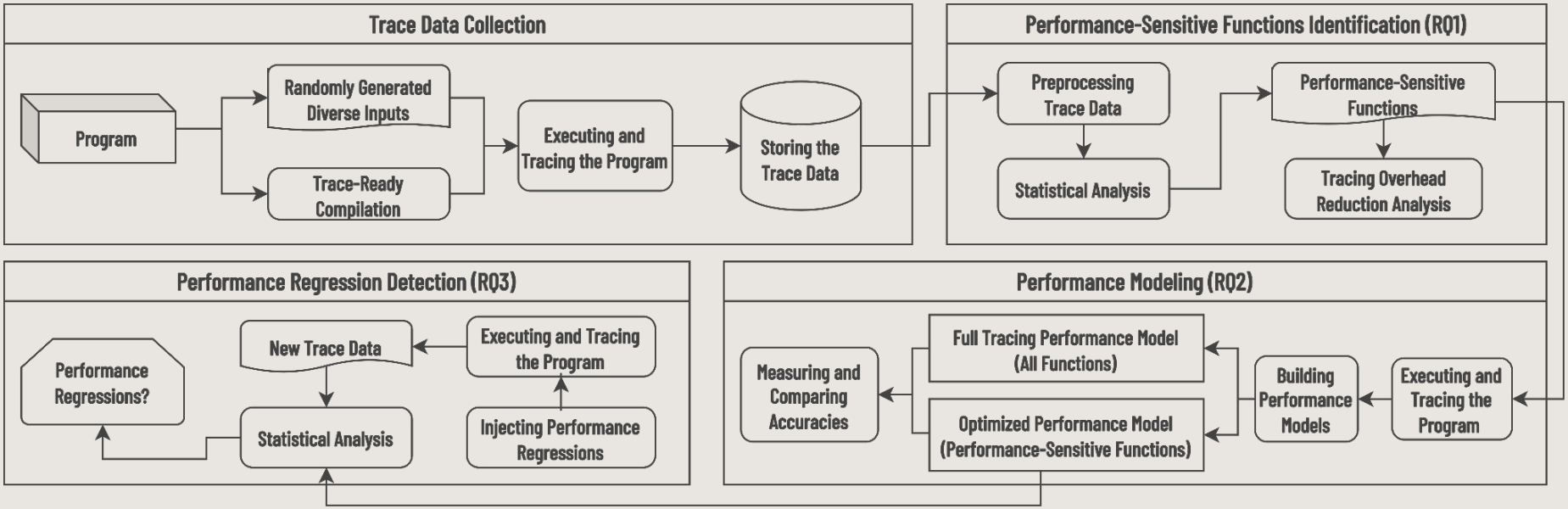
Methodology - Step 2



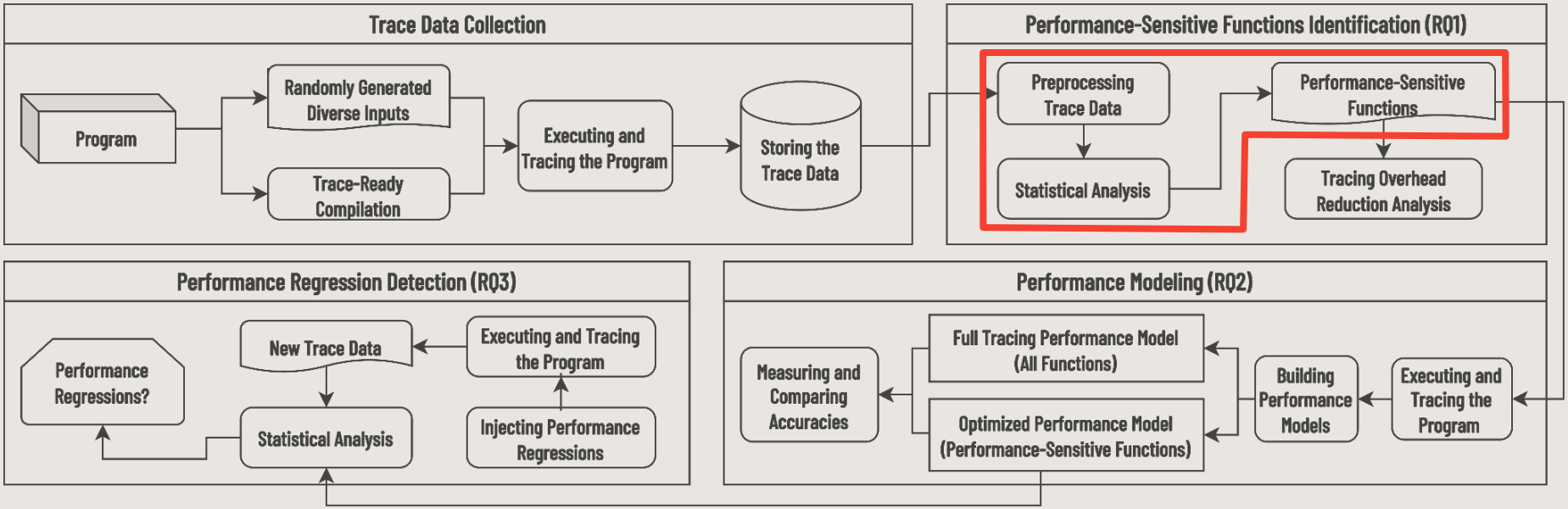
Methodology - Step 3



Methodology - Step 4



Methodology - Step 4



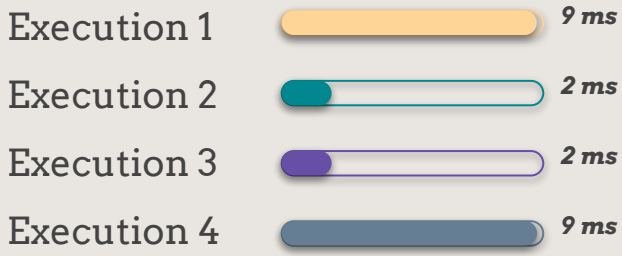
Statistical Analysis - Part 1

Shannon's Entropy

The Uncertainty or Randomness

Coefficient of Variation

The Fluctuation



Statistical Analysis - Part 1

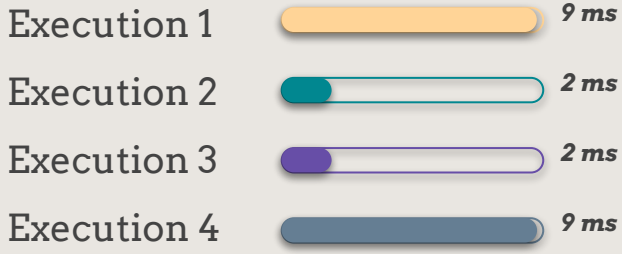
Shannon's Entropy

The Uncertainty or Randomness

***High** Entropy*
***Low** CoV*

Coefficient of Variation

The Fluctuation

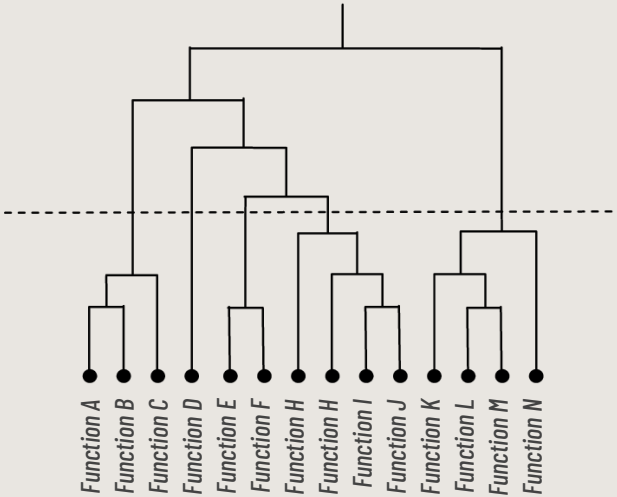


***High** CoV*
***Low** Entropy*

Statistical Analysis - Part 2

Performance Correlations

Highly Correlated Functions



Feature Significance

Contribution to Performance Model

- *Build a simple Linear Regression model from trace data*
- *Obtain the p-values of the model's coefficients*
- *Coefficients with p-value of 0.05 and less (i.e., p-value < 0.05)*

03

Evaluations

How much trace overhead did we reduce?
What is the accuracy of the performance models?
Can they detect performance regressions?

Trace Overhead Reduction

	Average Execution Time Overhead (Compared to Vanilla Execution in Percentage)						
Program	Full	Entropy		Coefficient of Variation		Performance Correlations	Feature Significance
		(w/o CR)	(w/ CR)	(w/o CR)	(w/ CR)		
SU2	**77.11%**	34.68%	12.52%	4.59%	**2.23%**	12.49%	3.95%
638.imagick_s	**168.37%**	100.35%	24.97%	28.3%	**14.29%**	28.29%	23.08%
631.deepsjeng_s	**471.88%**	31.34%	**0.45%**	56.96%	4.95%	11.47%	110.75%
	Average Storage Usage Overhead						
Program	Full	Entropy		Coefficient of Variation		Performance Correlations	Feature Significance
		(w/o CR)	(w/ CR)	(w/o CR)	(w/ CR)		
SU2	**748.58MB**	487.03MB	143.46MB	23.29MB	**7.45MB**	140.88MB	21.65MB
638.imagick_s	**899.36MB**	832.37MB	137.70MB	144.06MB	**30.16MB**	159.76MB	114.86MB
631.deepsjeng_s	**3.92GB**	285.35MB	**1.02MB**	515.92MB	34.65MB	81.78MB	976.85MB

Accuracy of Performance Models

Criterion		SU2							
		Linear Regression		Random Forest		CatBoost		XGBoost	
		MAE	R^2 Score	MAE	R^2 Score	MAE	R^2 Score	MAE	R^2 Score
Full		0.59	0.97	**0.22**	**0.99**	1.03	0.93	3.50	0.42
Entropy	w/o CR	0.65	0.97	0.26	**0.99**	0.96	0.93	2.83	0.52
	w/ CR	0.67	0.95	0.24	**0.99**	0.78	0.94	2.35	0.53
Coefficient of Variation	w/o CR	0.62	0.96	0.24	0.98	0.98	0.91	2.20	0.54
	w/ CR	2.27	0.58	0.69	0.91	1.70	0.75	2.27	0.46
Performance Correlations		0.69	0.95	0.41	0.97	0.81	0.93	2.35	0.53
Feature Significance		2.37	0.58	0.71	0.89	1.65	0.77	2.30	0.45

Regression Detections

The effectiveness of the optimized performance models in detecting performance regressions.
ES stands for *Effect Size*, which is calculated using *Cliff's Effect Size*.

	SU2					
Model	Without Regression (i.e., Baseline)		With Regression			
	P-Value	Effect Size	P-value > 0.05 or ES == N	ES==S	ES==M	ES==L
Random Forest	0.600	S. [0.202]	1/15	3/15	7/15	4/15
	631.deepsjeng_s					
Model	Without Regression (i.e., Baseline)		With Regression			
	P-Value	Effect Size	P-value > 0.05 or ES == N	ES==S	ES==M	ES==L
Linear Regression	0.510	N/A	2/15	1/15	1/15	11/15

04

Conclusion

Everything in a nutshell

The Trade-Off...

Performance Model's Accuracy

Fully tracing may result in very accurate performance models

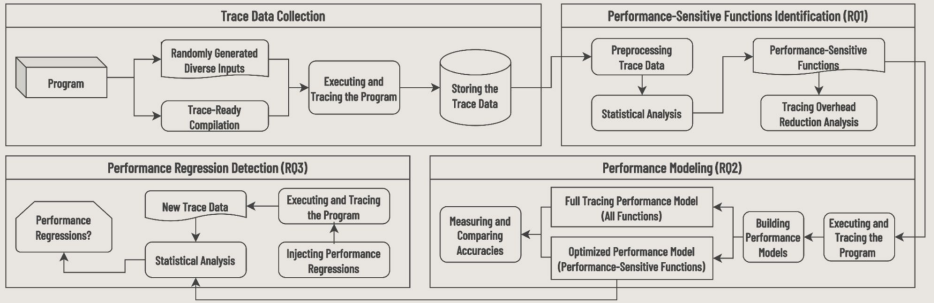
- R^2 Score > 95%
- Mean Prediction Error < 5%

Tracing Overhead

Tracing of high-computational applications is quite expensive

- Mean Execution Time Overhead >> 50%
- Mean Storage Usage Overhead >> 1000%

Methodology - Step 4



Statistical Analysis - Part 1

Shannon's Entropy

The Uncertainty or Randomness

Coefficient of Variation

The Fluctuation

Trace Overhead Reduction

Program	Average Execution Time Overhead (Compared to Vanilla Execution in Percentage)						
	Full	Entropy		Coefficient of Variation		Performance Correlations	Feature Significance
		(w/o CR)	(w/ CR)	(w/o CR)	(w/ CR)		
SU2	**77.11%**	34.68%	12.52%	4.59%	**2.23%**	12.49%	3.95%
638.imagick_s	**168.37%**	100.35%	24.97%	28.3%	**14.29%**	28.29%	23.08%
631.deepsjeng_s	**471.88%**	31.34%	**0.45%**	56.96%	4.95%	11.47%	110.75%

Program	Average Storage Usage Overhead						
	Full	Entropy		Coefficient of Variation		Performance Correlations	Feature Significance
		(w/o CR)	(w/ CR)	(w/o CR)	(w/ CR)		
SU2	**748.58MB**	487.03MB	143.46MB	23.29MB	**7.45MB**	140.88MB	21.65MB
638.imagick_s	**899.36MB**	832.37MB	137.70MB	144.06MB	**30.16MB**	159.76MB	114.86MB
631.deepsjeng_s	**3.92GB**	285.35MB	**1.02MB**	515.92MB	34.65MB	81.78MB	976.85MB

THANKS!

kaveh.shahedi@polymtl.ca

linkedin.com/in/kavehshahedi