A Framework for Cost-effective Dependence-Based Dynamic Impact Analysis

Haipeng Cai and Raul Santelices

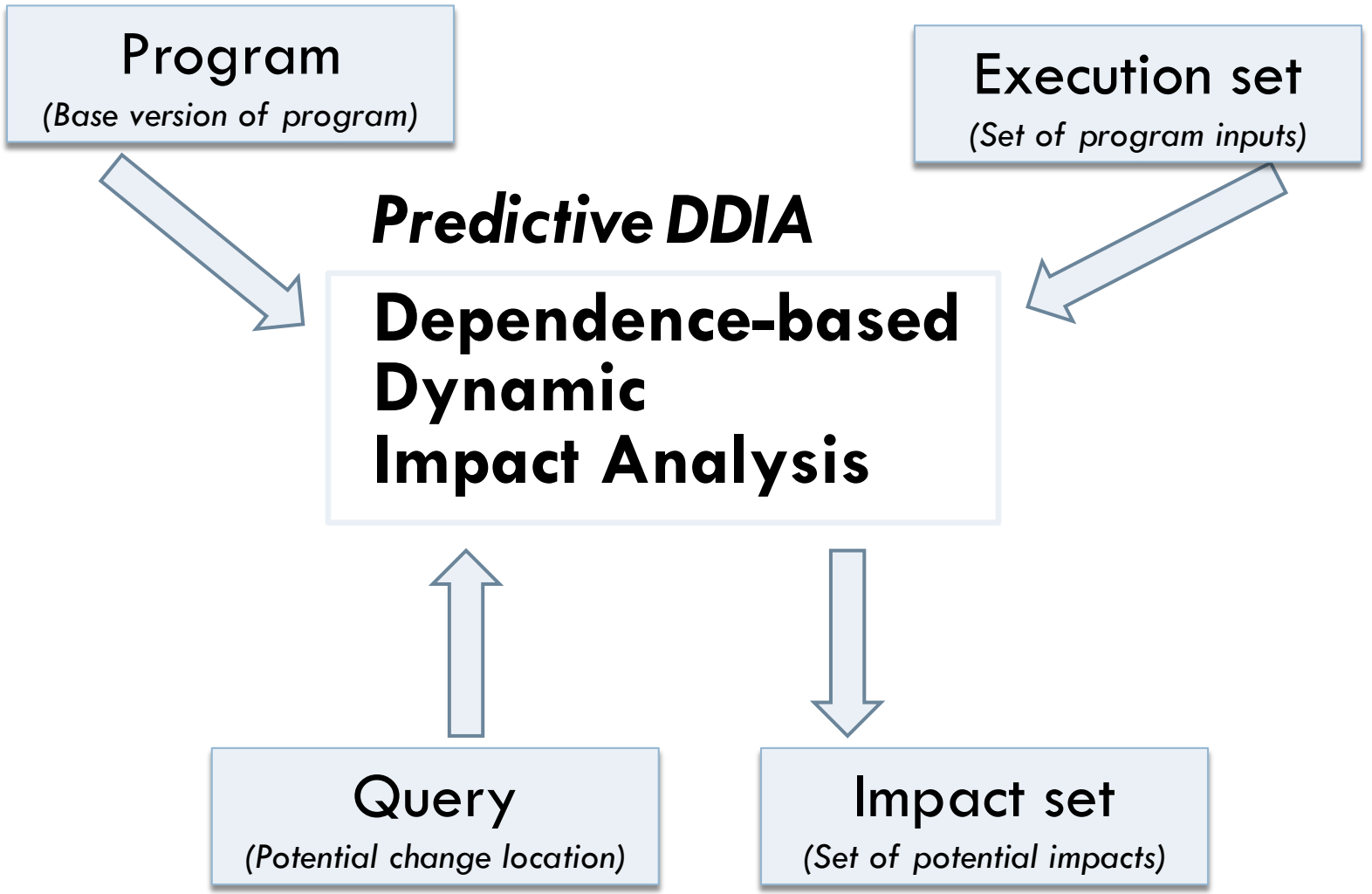
Department of Computer Science and Engineering

University of Notre Dame

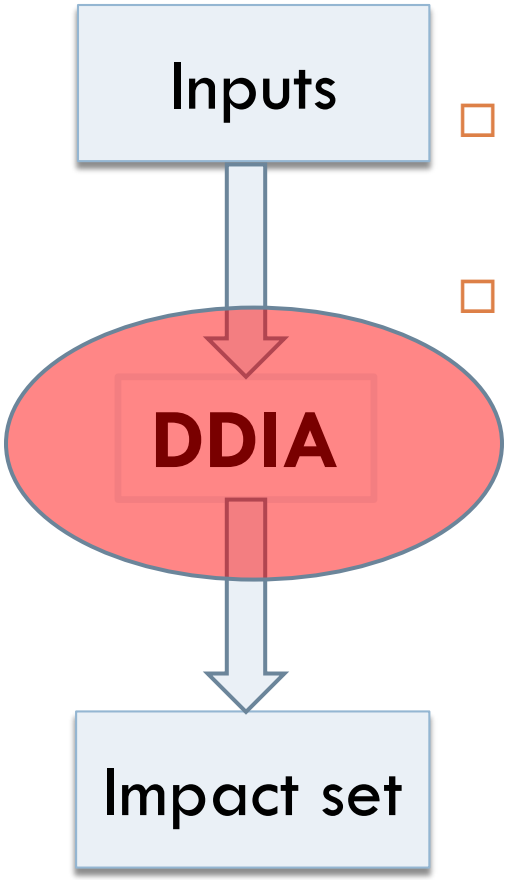
SANER 2015

Supported by ONR Award N000141410037

Background

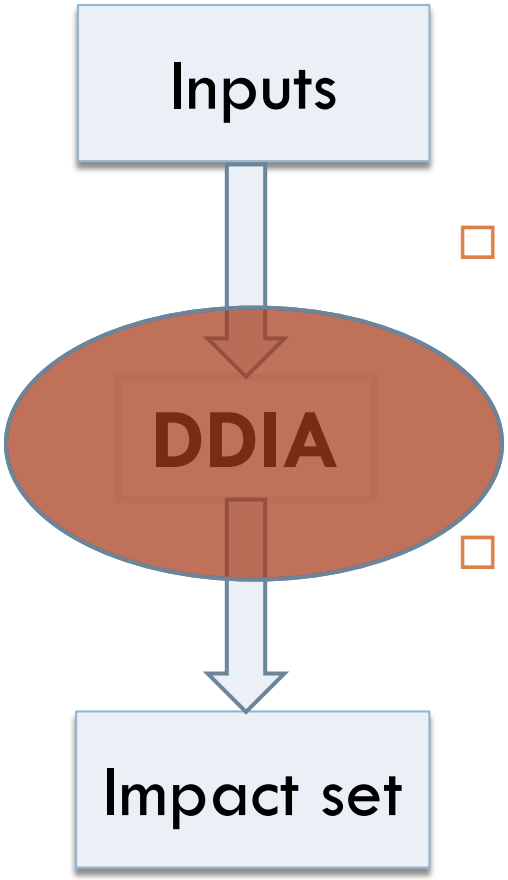

Program
(Base version of program)
Execution set
(Set of program inputs)
Predictive DDIA
Dependence-based
Dynamic
Impact Analysis
Query
(Potential change location)
Impact set
(Set of potential impacts)

Problem



- Efficient approaches are too imprecise
  (e.g., PathImpact/EAS [T. Apiwattanapong et al., 2005])
- Precise approaches are too expensive
  (e.g., dynamic slicing [X. Zhang et al., 2004])
- Developers need techniques of multiple levels of cost-effectiveness tradeoffs for diverse needs
  (e.g., budgets versus the level of precision needed) [C.R. Souza et al., 2008]

Approach

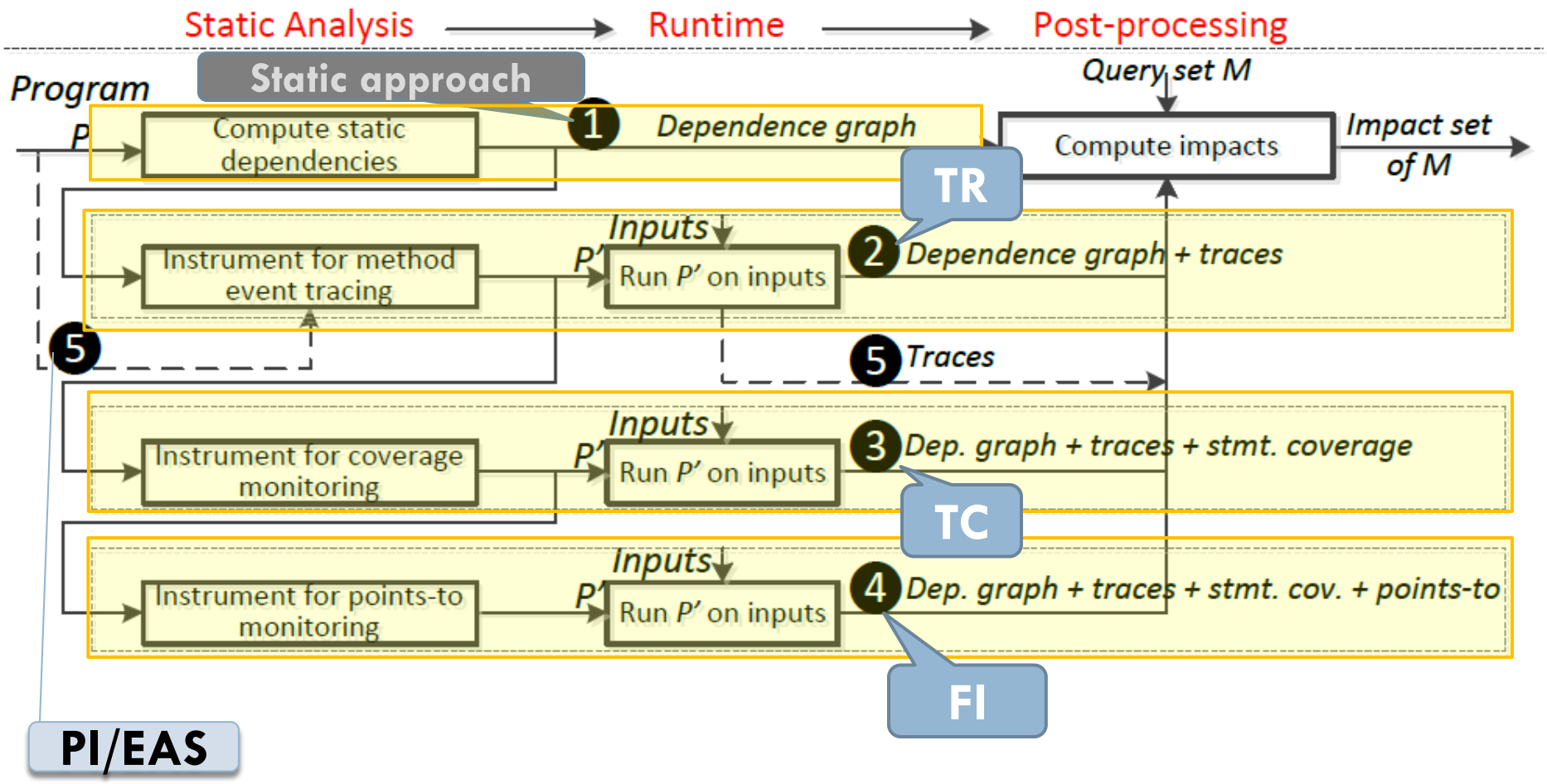


- Utilize static dependencies in collaboration with method-level execution traces (i.e., hybrid approach)
- Exploit additional dynamic information
  - Statement coverage
  - Dynamic points-to data [M. Mock et al., 2005]
- Guide trace-based impact computation with both static and dynamic information

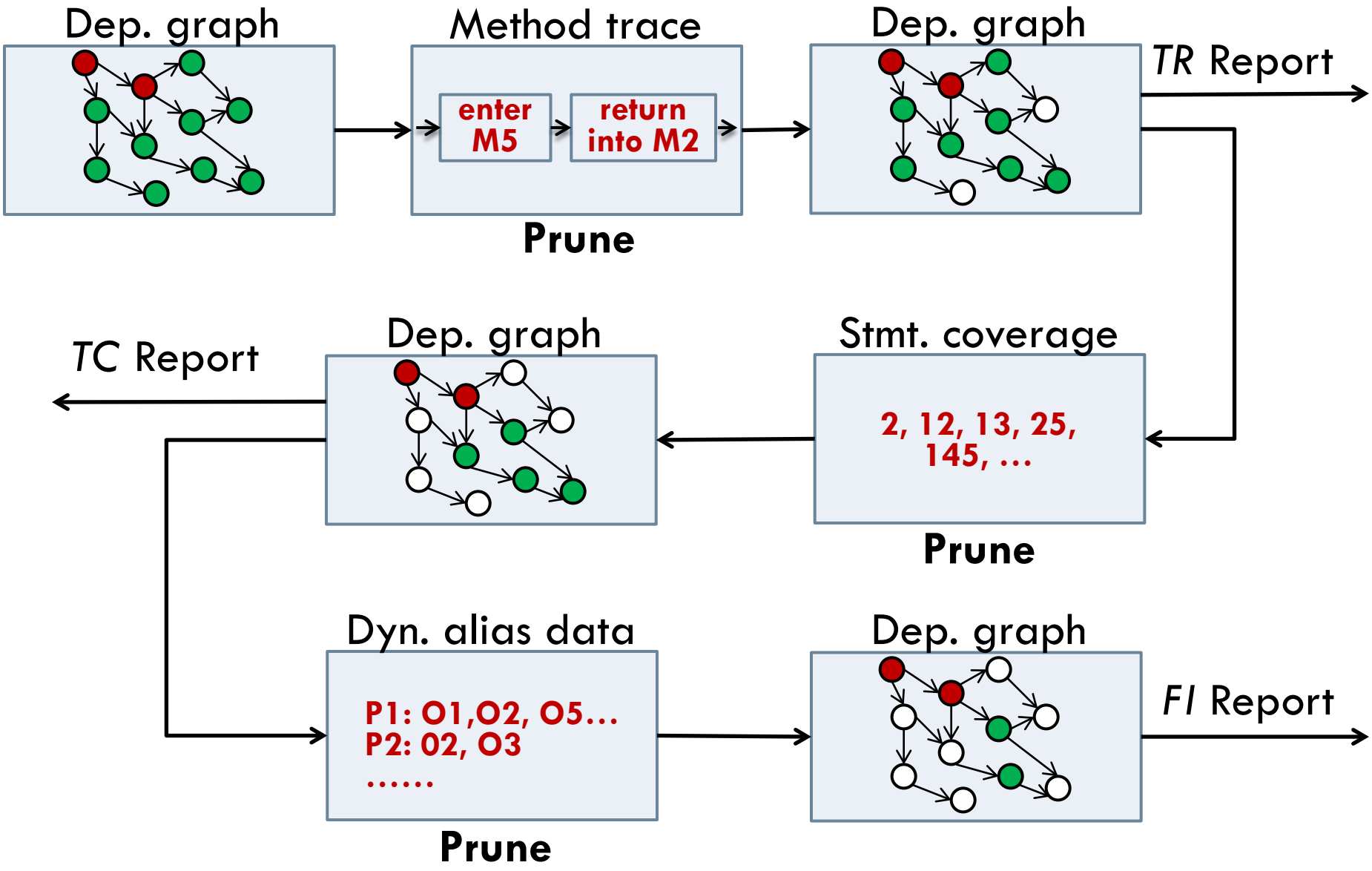
Solution

- A framework that unifies analysis techniques of various cost-effectiveness tradeoffs
  - Including existing representative options (PI/EAS)
  - Spawning three new instances
- Three new instances
  - TR: static dependencies + method TRaces
  - TC: TR + statement Coverage
  - FI: Full Information -- TC + dynamic points-to data

The Framework



Algorithm


Dep. graph
Method trace
enter M5
return into M2
Prune
Dep. graph
TR Report
TC Report
Dep. graph
Stmt. coverage
2, 12, 13, 25, 145, …
Prune
Dyn. alias data
P1: O1,O2, O5…
P2: 02, O3
……
Prune
Dep. graph
FI Report

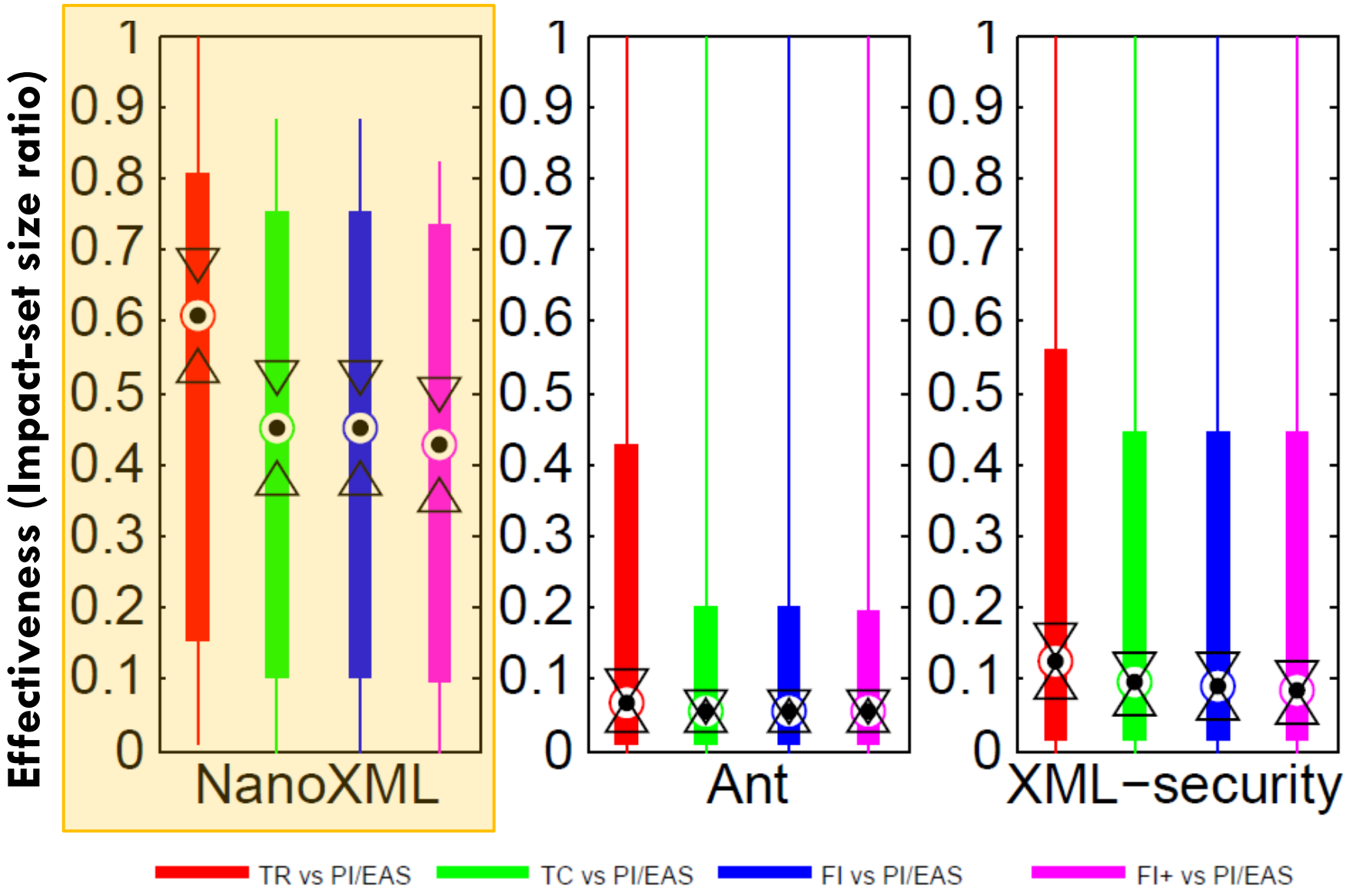
Experimental setup

- Subjects
  - 7 Java programs
  - Up to 212 KLOC in size (1k ~ 100k)
- Techniques
  - PI/EAS (baseline), TR, TC, FI (, FI+)
- Metrics
  - Effectiveness
    - Impact-set size ratios to baseline
  - Cost
    - Computation time
    - Storage space
  - Average cost-effectiveness
    - $$\frac{\text{Percentage of impact-set reduction}}{\text{factor of time cost increase}}$$

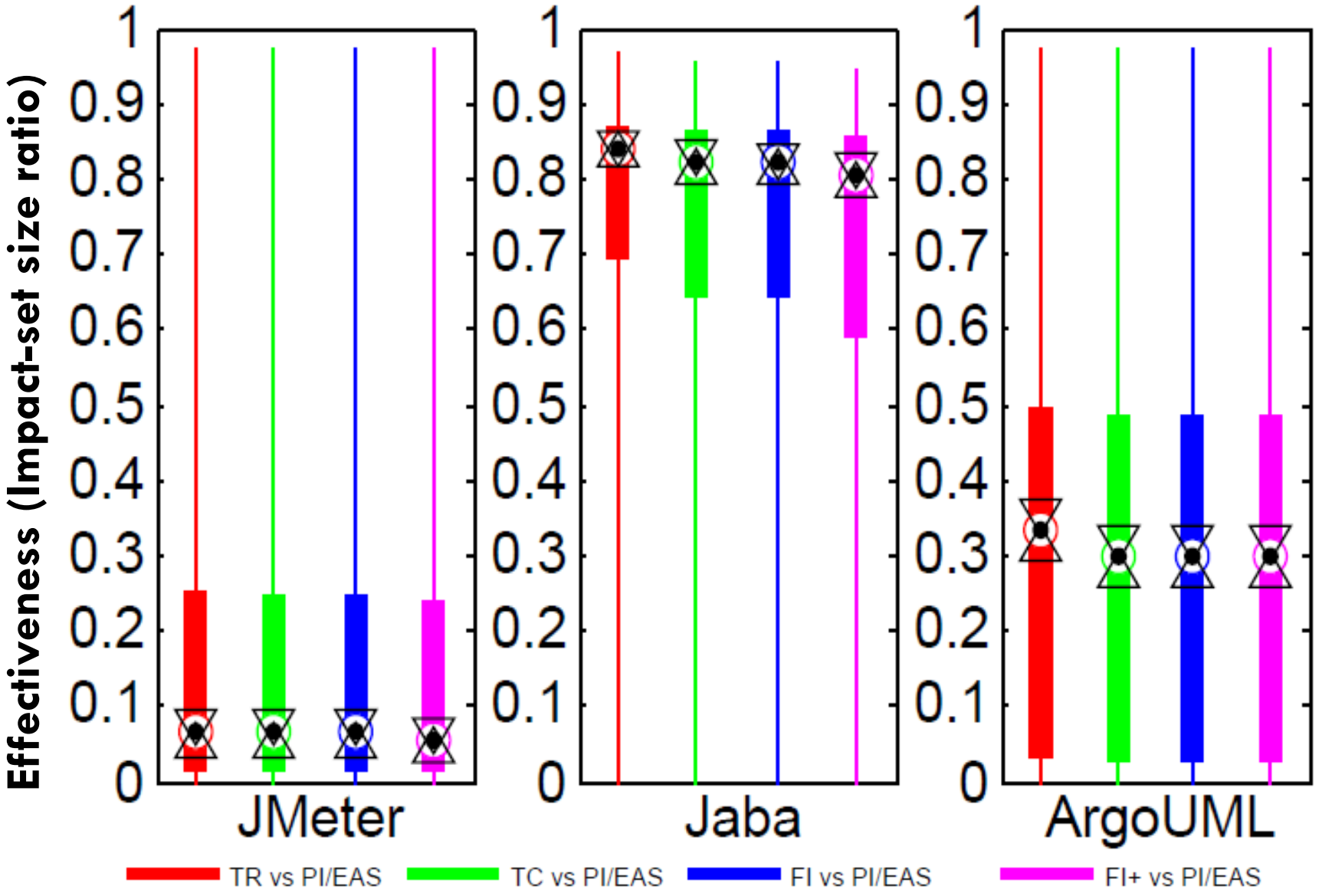
Research questions

- How do the techniques compare in terms of effectiveness?
- How do the techniques compare in terms of costs?
- What are the effects of different forms of dynamic data on the DDIA cost-effectiveness?

Result: effectiveness



Result: effectiveness



Research questions

- How do the techniques compare in terms of effectiveness?
- How do the techniques compare in terms of costs?
- What are the effects of different forms of dynamic data on the DDIA cost-effectiveness?

Result: querying cost

Subject	PI/EAS (seconds)	Query time of our techniques (seconds)			
		TR	TC	FI	FI+
Schedule1	0.70	14.60	15.72	19.24	44.26
NanoXML	0.07	6.24	6.35	5.60	7.97
XML-security	0.04	7.43	8.01	8.15	16.89
JMeter	0.02	2.25	2.30	1.82	2.18
Ant-v0	0.05	3.19	3.39	3.31	5.24
Jaba	0.29	78.34	99.68	82.55	105.18
ArgoUML	0.05	15.95	15.98	12.60	15.82
Overall	0.11	26.33	31.96	26.62	35.04

Result: other costs

- Static-analysis costs in seconds

Subject	PI/EAS	TR	TC	FI/FI+
Schedule1	5	6	11	17
NanoXML	11	14	25	39
Ant	27	142	170	311
XML-security	33	158	190	280
JMeter	38	372	408	764
Jaba	55	289	326	600
ArgoUML	172	7,465	7,542	11,998
Overall	73	2,047	2,115	3,392

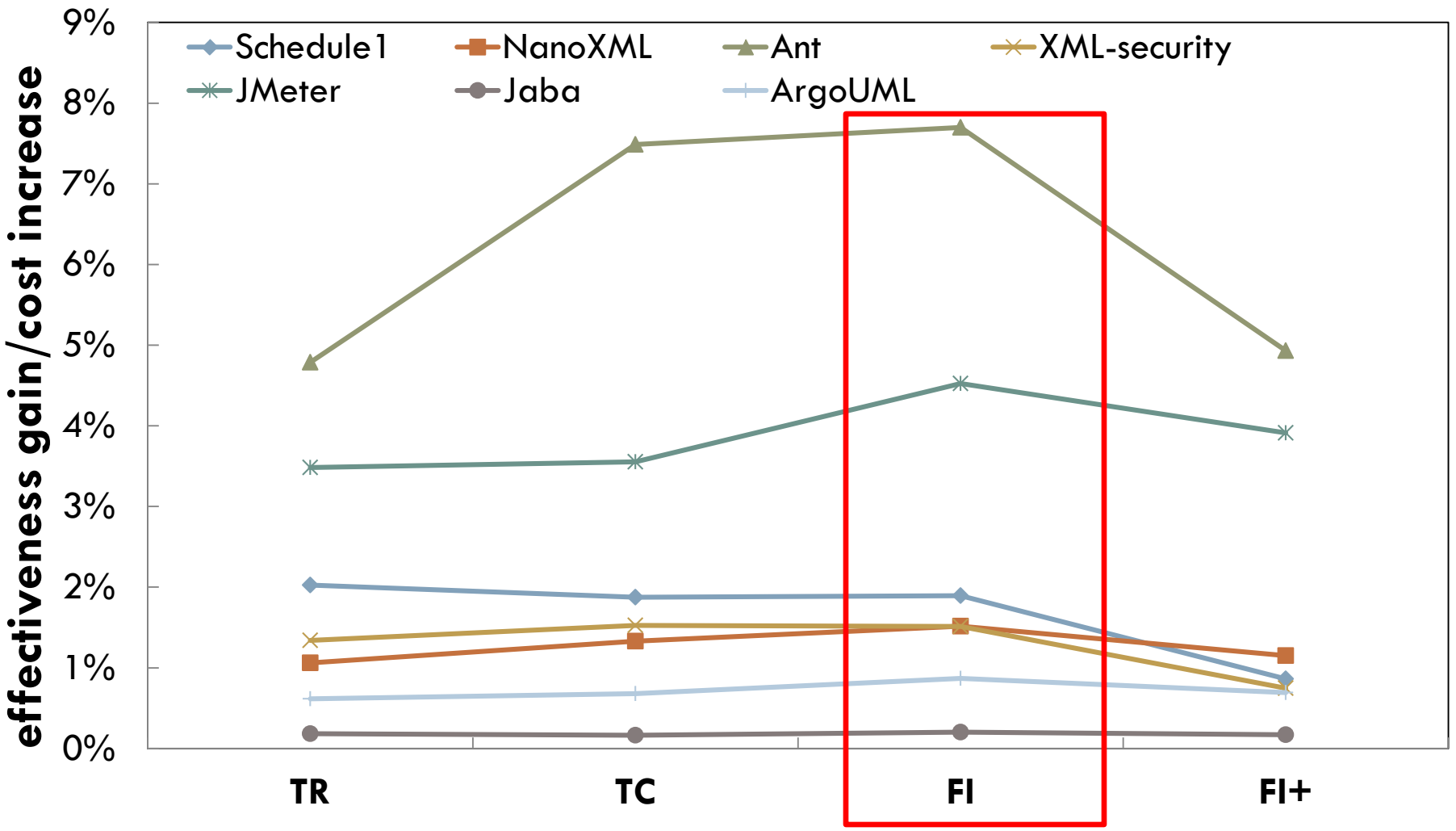
- Runtime costs: < 1m
- Space costs: < 4MB

Research questions

- How do the techniques compare in terms of effectiveness?
- How do the techniques compare in terms of costs?
- What are the effects of different forms of dynamic data on the DDIA cost-effectiveness?

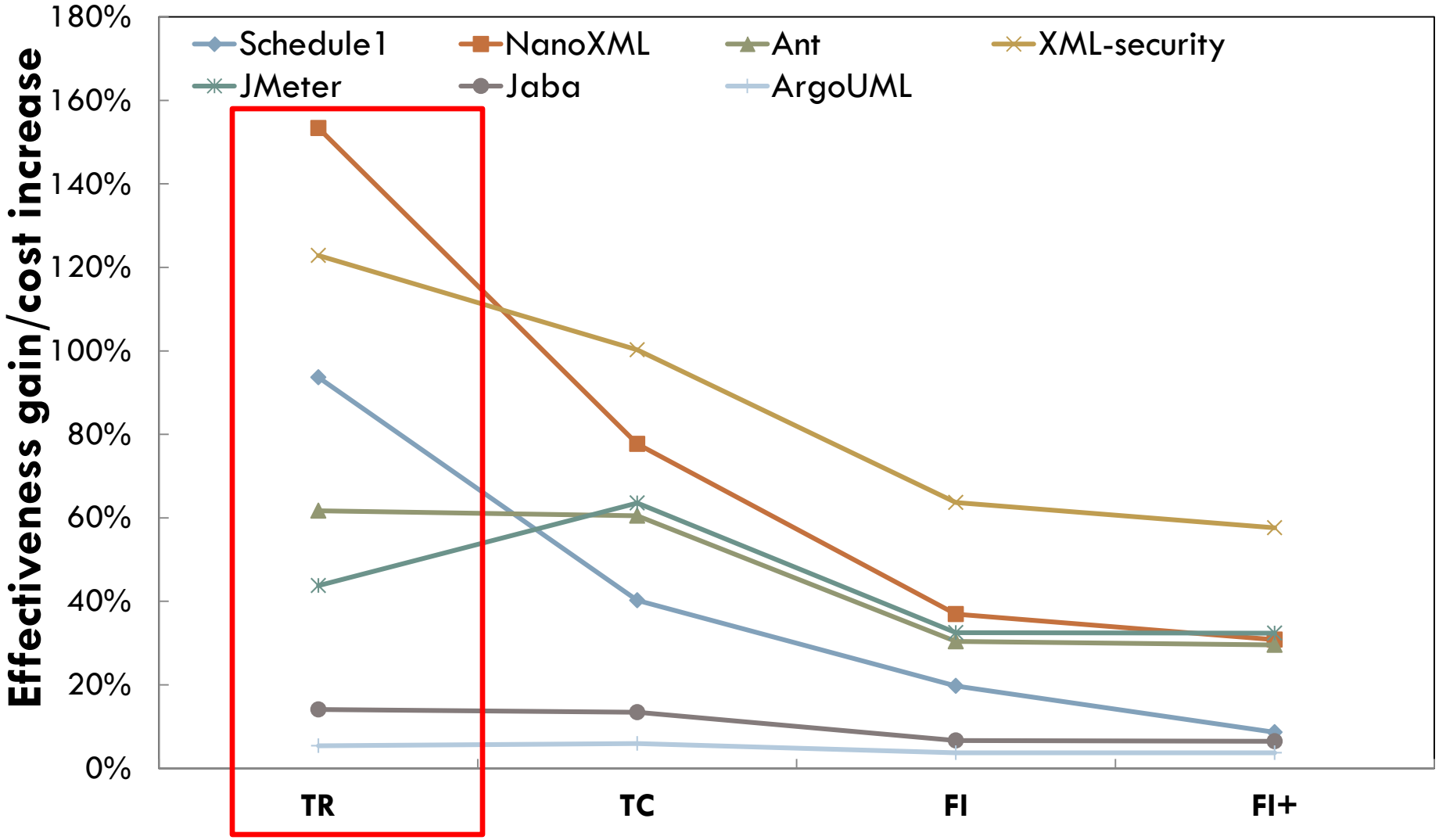
Result: cost-effectiveness

- With respect to querying costs



Result: cost-effectiveness

- With respect to other costs



Conclusions

- A framework that unifies existing and new DDIA techniques, and offers multiple-level cost-effectiveness options
- New techniques greatly reducing impact-set sizes, implying large improvement in precision
- Statement coverage has generally stronger effects on DDIA cost-effectiveness than dynamic points-to data

Acknowledgements

Office of Naval Research for funding

All of you for time and attention

Q&A

The proposed framework offers multiple-level trade-offs between cost and effectiveness of dynamic impact analysis.

Haipeng Cai

http://cse.nd.edu/~hcai/

hcai@nd.edu

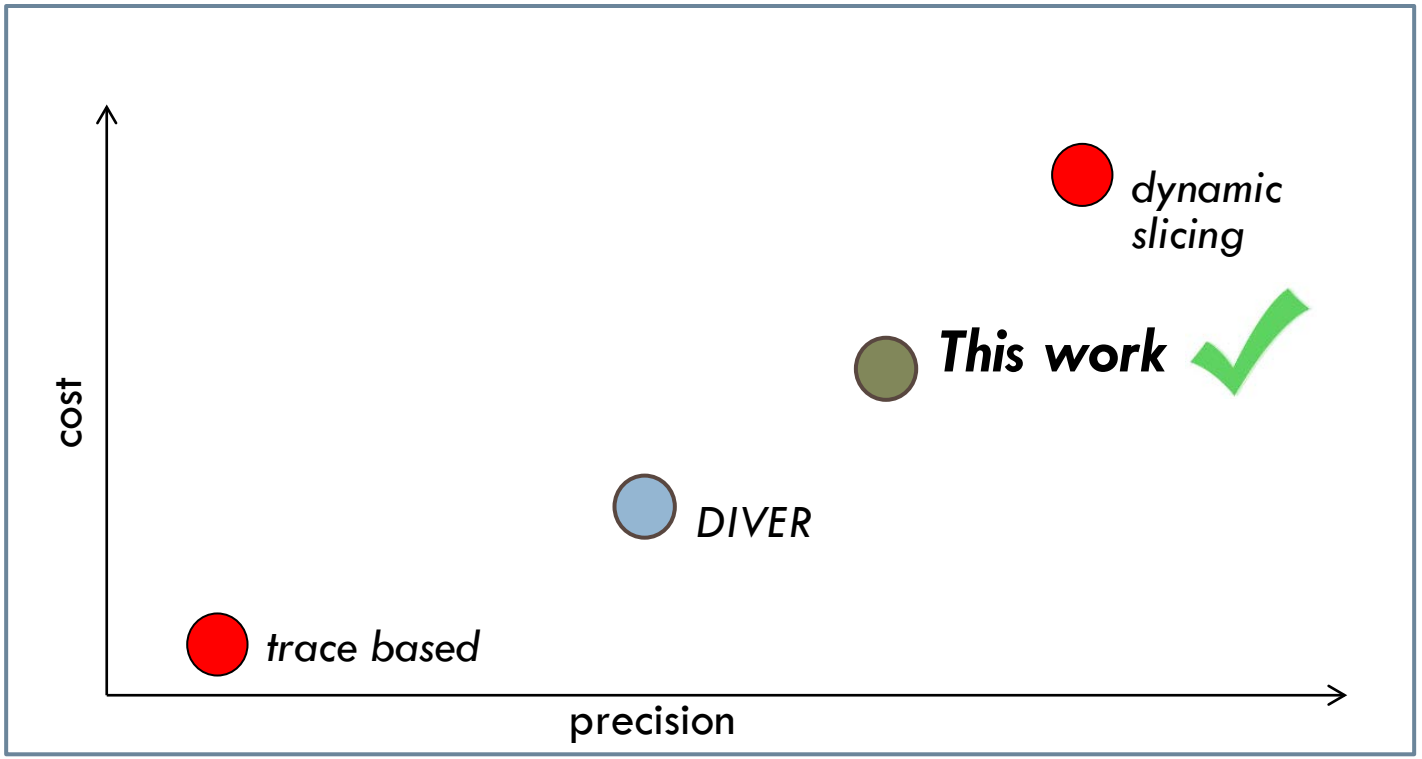
Subject programs

Subject	KLOC	#Methods	#Tests
Schedule1	0.3	24	2,650
NanoXml	3.5	282	214
Ant-v0	18.8	1,863	112
XML-security-v1	22.4	1,928	92
JMeter-v2	35.5	3,054	79
Jaba	37.9	3,332	70
ArgoUML-r3121	102.4	8,856	211

Controversial/provocative statement

- Achieving 100% recall *with respect to actual impacts* for dynamic dependence analysis is impossible.
- Impact analysis is being emphasized all the time but practitioners mostly still stick to old-fashioned ways relying on manual efforts, what are possible obstacles there?

Design space of cost-effective DDIA



Key idea:

Incrementally **prune** methods NOT dependent on the query