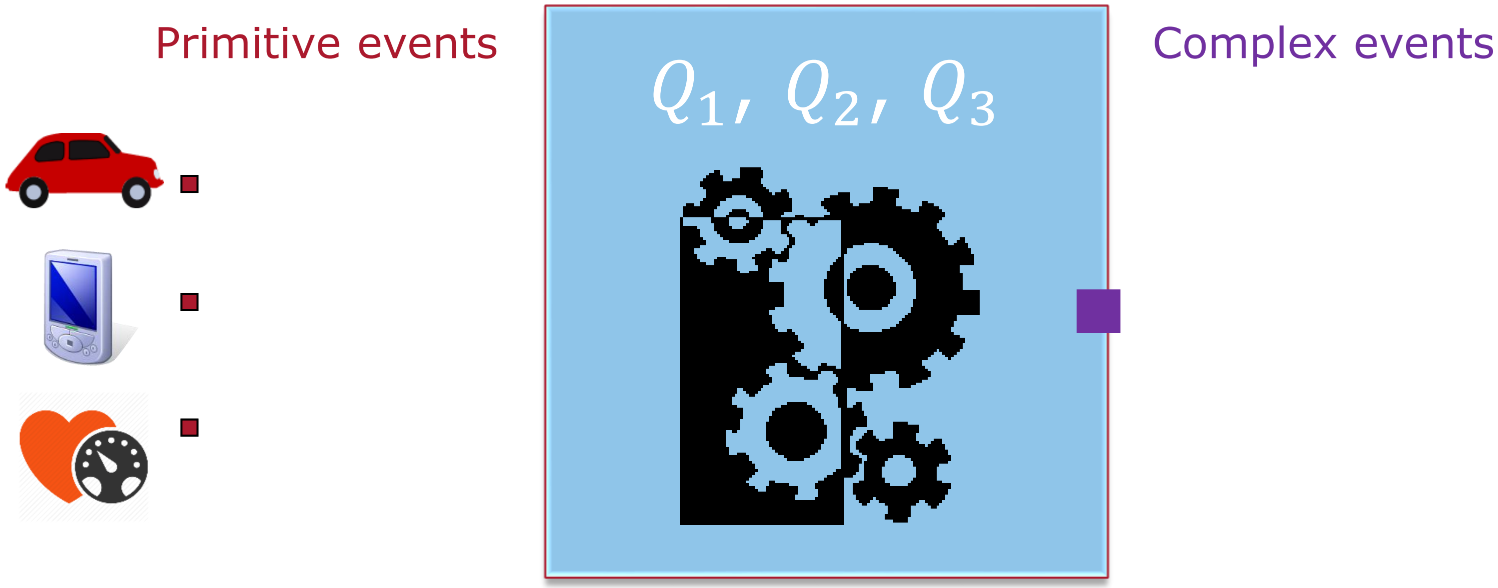
CAESAR:
Context-Aware Event Stream Analytics in Real time

Olga Poppe, Chuan Lei,
Elke A. Rundensteiner, and Dan Dougherty

March 18, 2016

Complex Event Processing

CEP engine



The **same** workload of **independent** event queries is **continuously** evaluated

Application Context

- Event compositions signify application contexts
- Most event queries are appropriate only in certain contexts
- They can be safely suspended otherwise

Examples of application contexts:

- **Emergency management:** normal, crowded, fire
- **Health care:** safe, warning, violation
- **Algorithmic trading:** hold, buy, sell
- **Financial fraud:** approved, suspicious, fraud

Traffic Management Use Case

- **140 hours idling in traffic** due to congestion in 10-worst U.S. traffic corridors per year [The Wall Street Journal]
- **Health cost of $18 billion** due to traffic noise and pollution in the USA's 83 largest urban areas in 2010 [USA Today]
- **1.24 million deaths** due to traffic injuries worldwide in 2010 [Wikipedia]

Traffic Management Contexts

Accident	Congestion	Clear
		
Accident warning	Toll notification	Statistics
Route re-computation	Route re-computation	Local services

Goal is to leverage application contexts to speed up system responsiveness

Challenges

- Rich semantics
  - Complex conditions implying a context
  - Unknown and unbounded context duration
  - Multiple inter-dependent event queries
- Readable specification
- Real time responsiveness

State-of-the-art Approaches

	CEP Systems (Esper, StreamInsight)	CAESAR	Business Models (BPMN, UML)
Expressive event queries	✓	✓	
Application contexts		✓	✓
Context-aware optimizations		✓	

Contributions & Outline

CAESAR system:

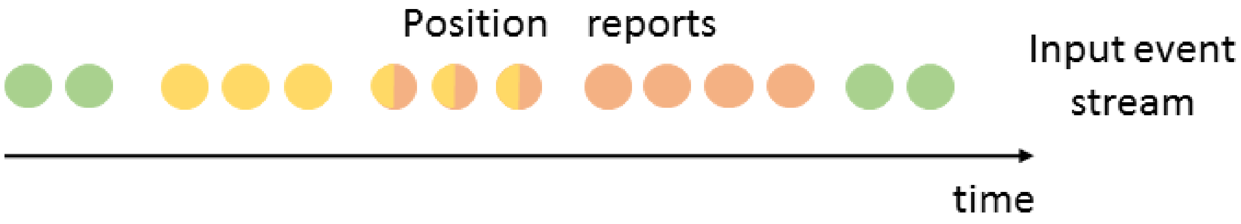
- Graphical model
- Context-aware algebra
- Context-driven optimization techniques
- Execution infrastructure

Performance evaluation

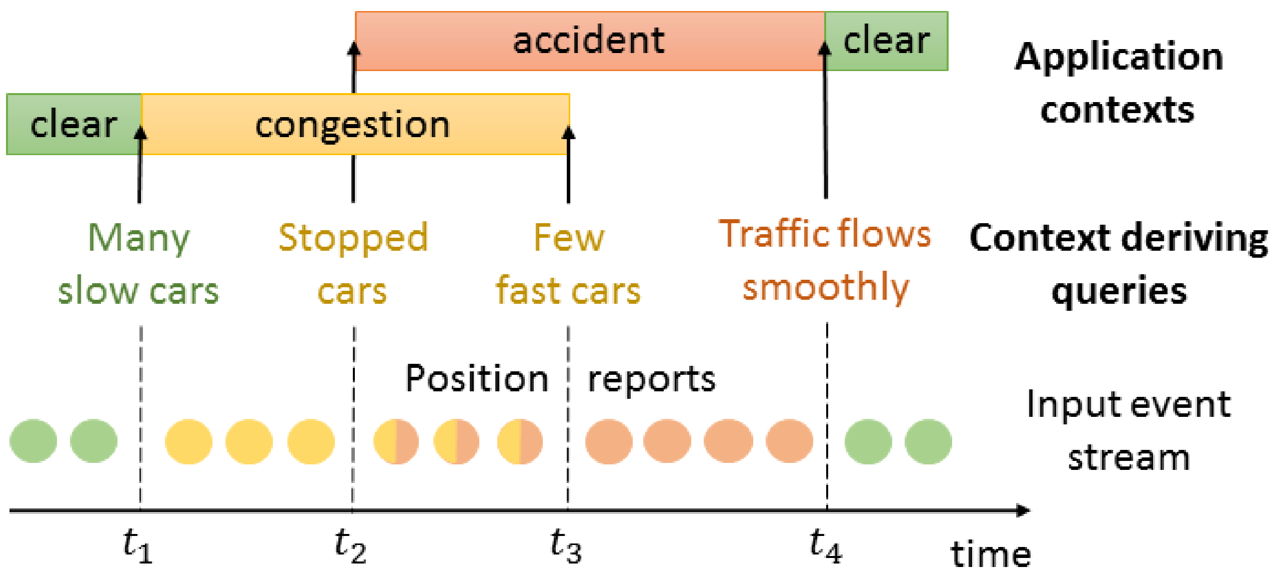
Outline

CAESAR Model

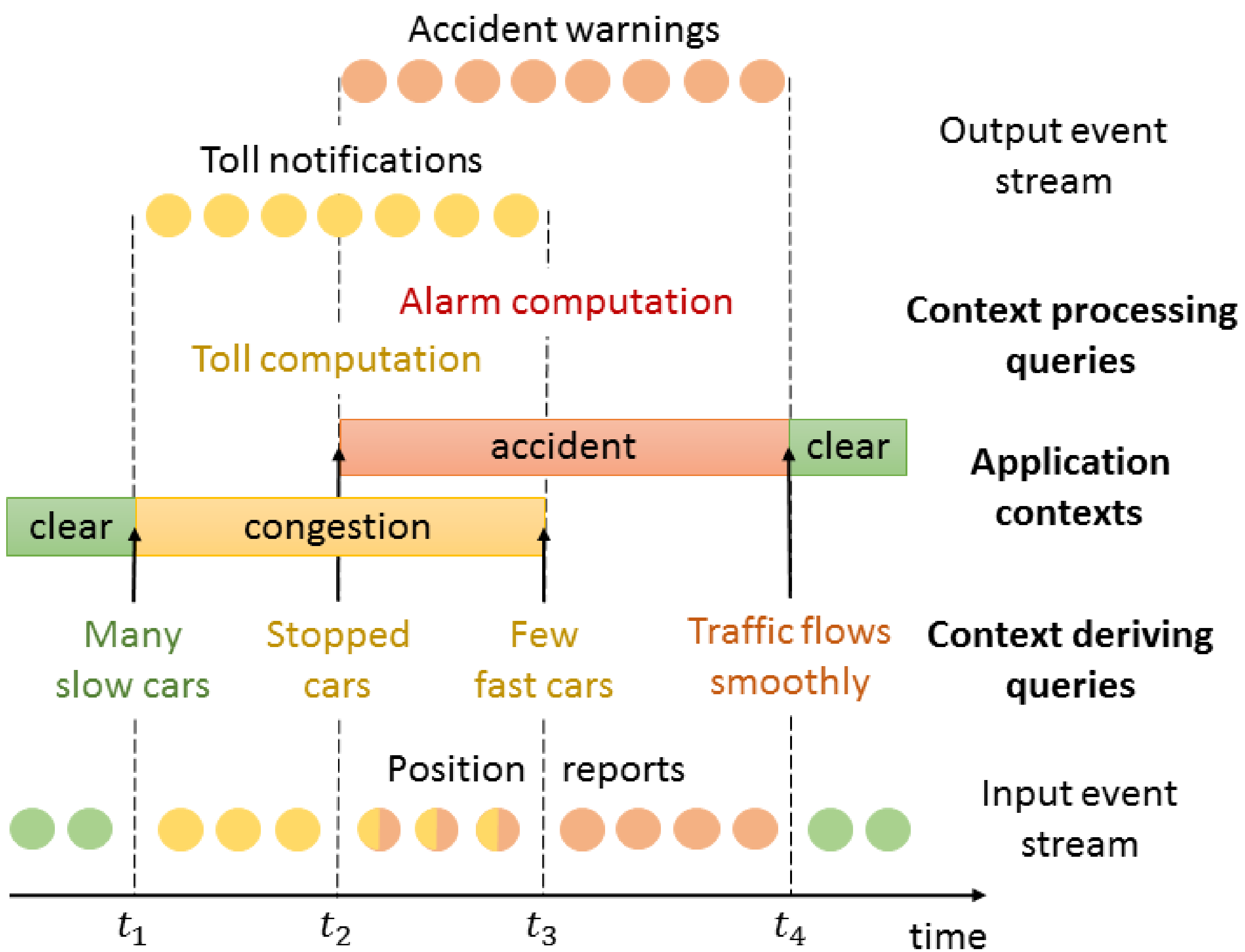
Context-aware Event Stream Analytics



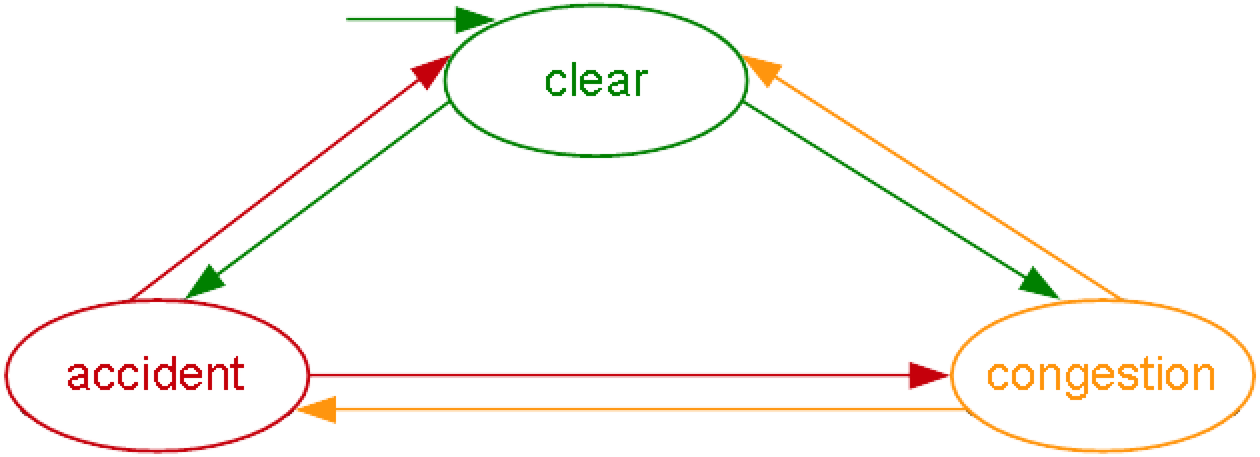
Context-aware Event Stream Analytics



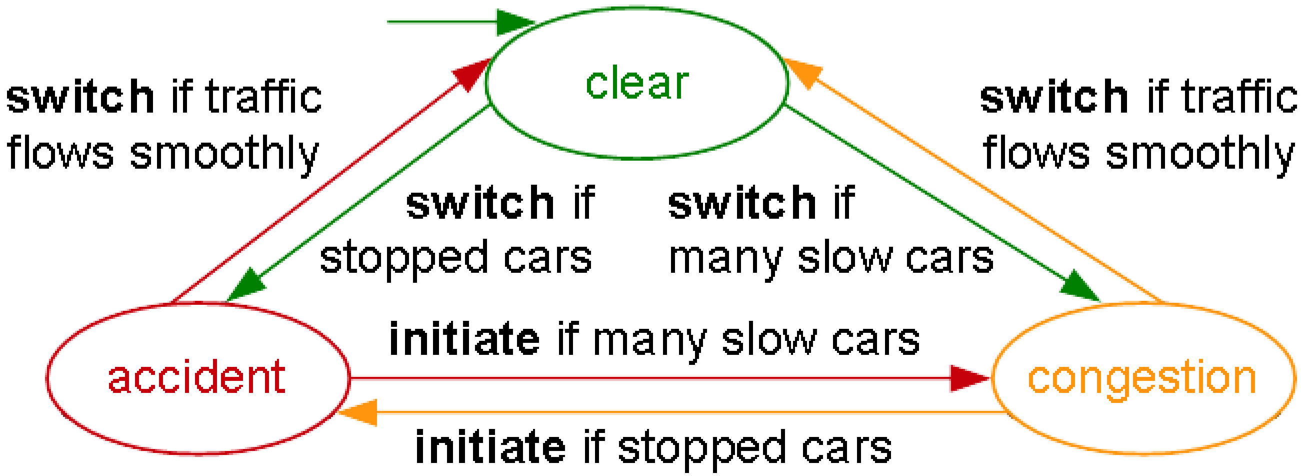
Context-aware Event Stream Analytics



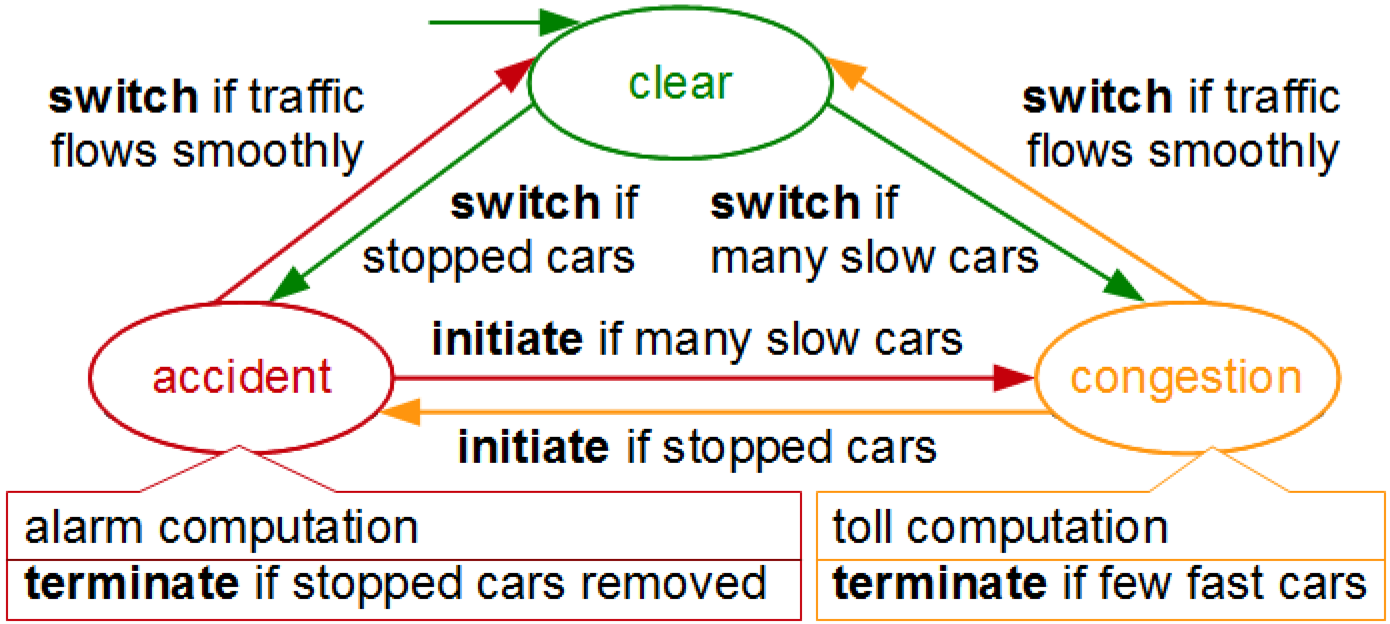
Application Contexts



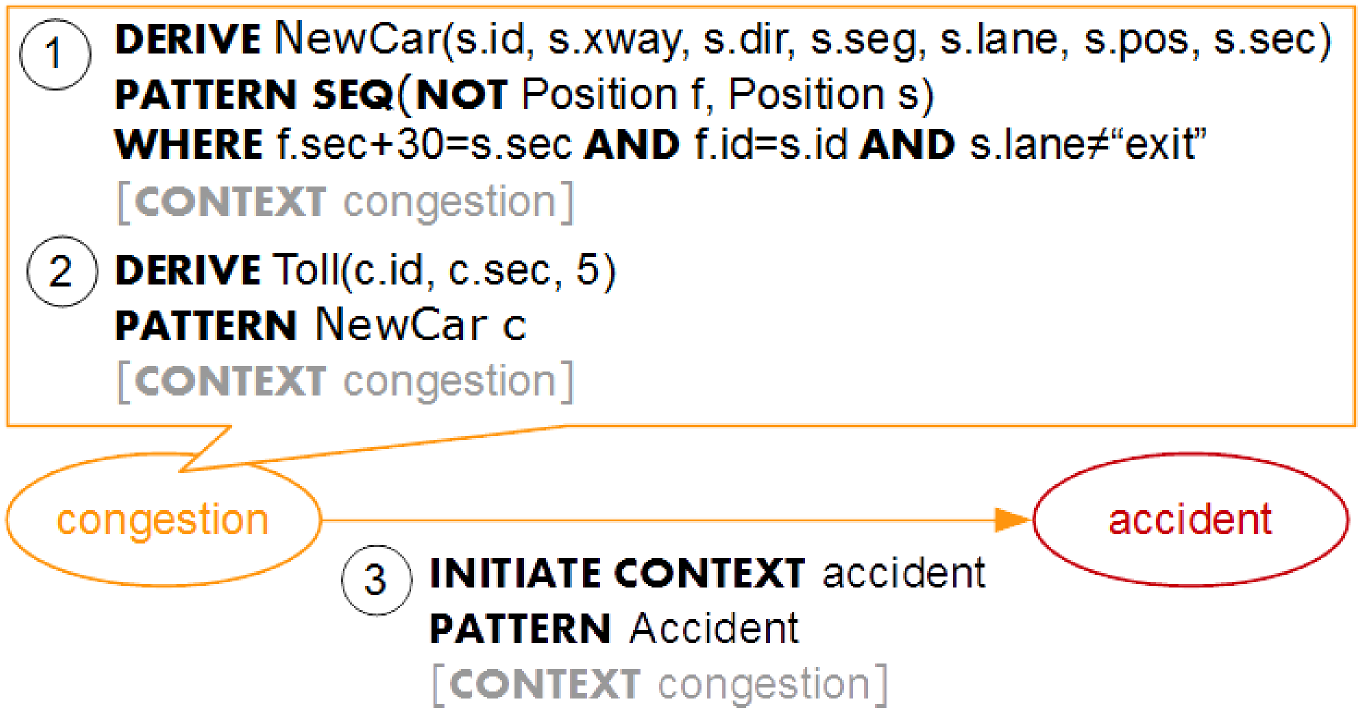
Context Deriving Queries



Context Processing Queries



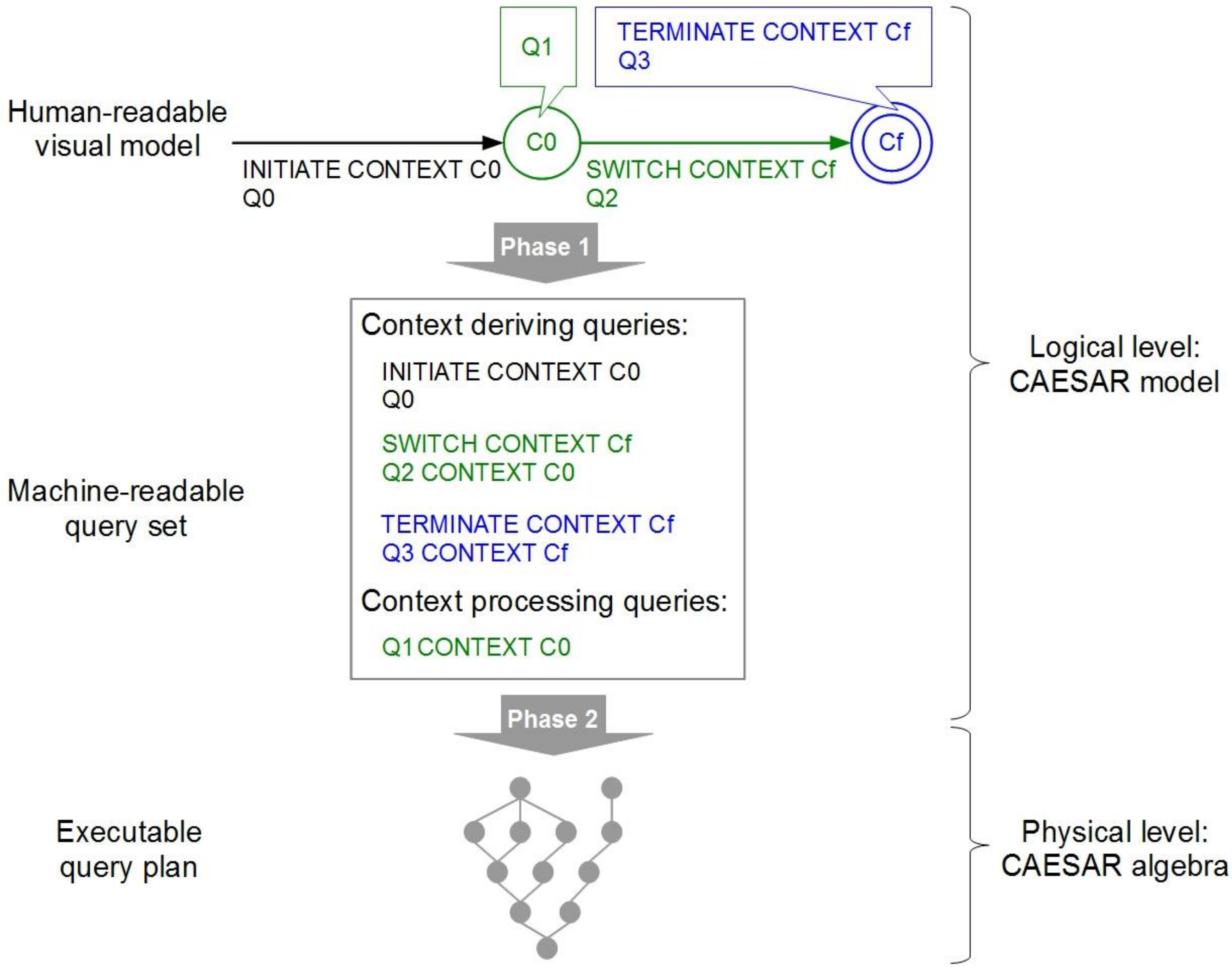
Context-aware Event Queries



Outline

CAESAR Algebra

Context-preserving Plan Generation



CAESAR Algebra Operators

1. Context initiation $CI_{\mathrm{c}}(I, W)$
2. Context termination $CT_{\mathrm{c}}(I, W)$
3. Context window $CW_c(I, W)$
4. Filter $FI_\theta(I)$
5. Projection $PR_{A,E}(I)$
6. Event pattern $P(I)$

Runtime Context Maintenance

Context bit vector W:

0	1	0	0	1	0	0	0	0	0

Context types: c_a, c_b, … c_z

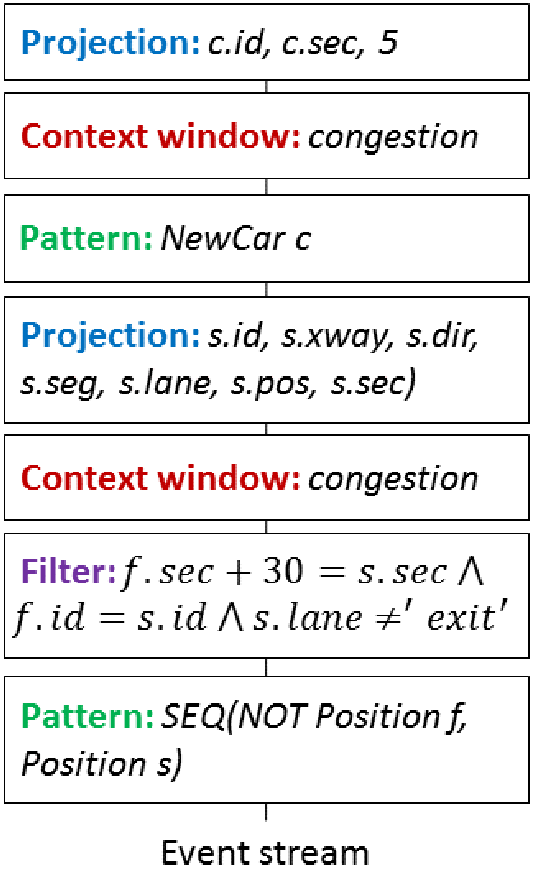
Time stamp $W.time$

- Updated by the context initiation & termination operators
- Accessed by the context window operator
- Synchronized by the time driven scheduler

Translation from Query Set to Algebra Plan

DERIVE Toll(c.id, c.sec, 5)
PATTERN NewCar c
CONTEXT congestion

DERIVE NewCar(s.id, s.xway, s.dir,
s.seg, s.lane, s.pos, s.lane)
PATTERN **SEQ**(**NOT** Position f, Position s)
WHERE f.sec+30=s.sec **AND**
f.id=s.id **AND**
f.lane≠' exit'
CONTEXT congestion



Outline

CAESAR Optimizer

CAESAR Optimizer Overview

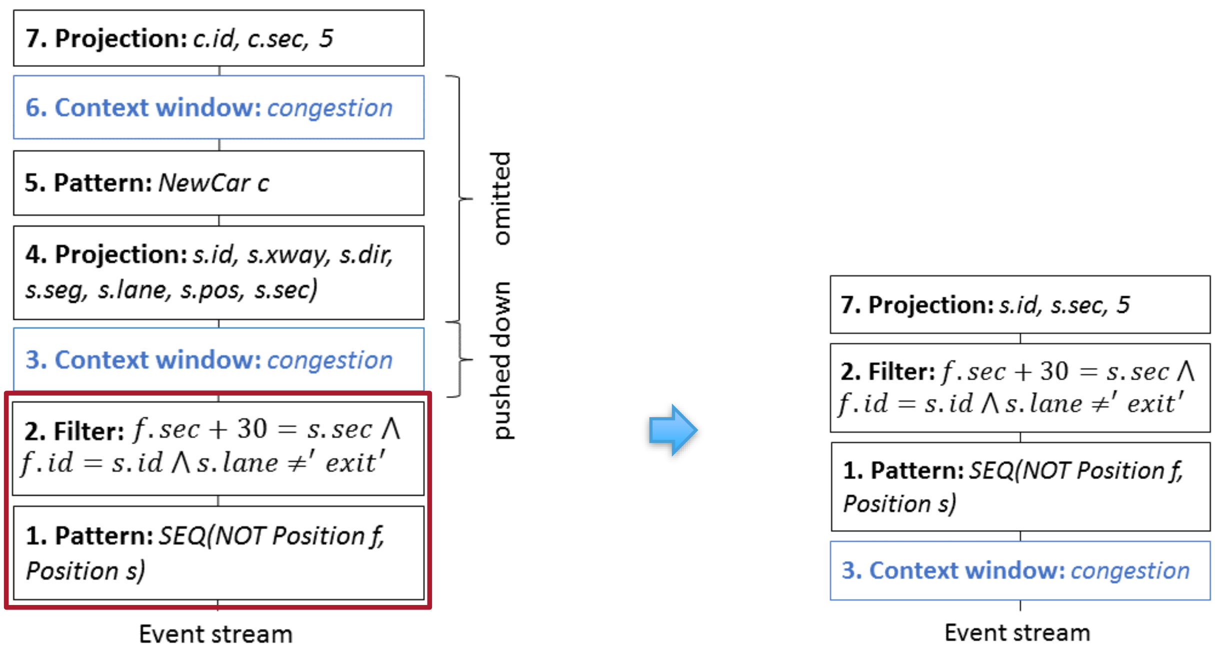
Problem statement:

Given a workload of **context-aware event queries,** our optimization problem is to find an optimized query plan for this workload with **minimal CPU cost.**

Context-aware optimization techniques:

- Context window push down strategy
- Context workload sharing algorithm

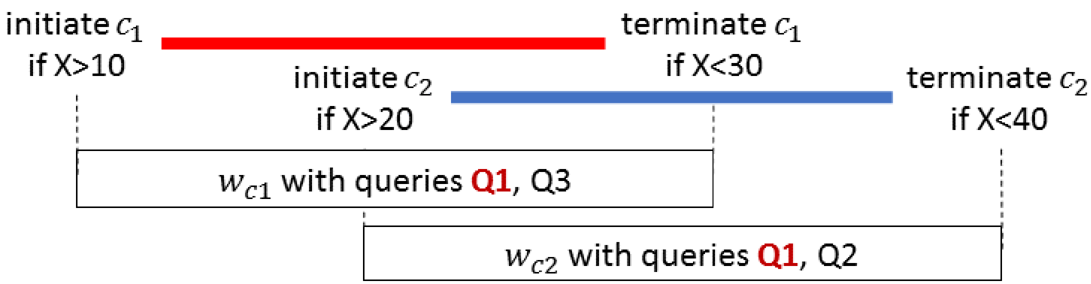
Context Window Push Down Strategy



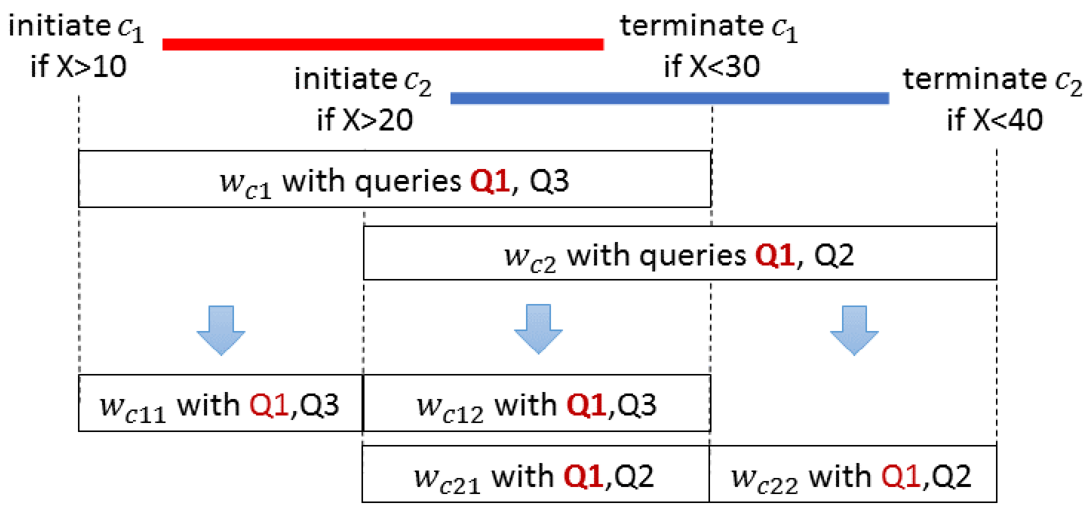
Performance benefits:

- Suspension of irrelevant operators
- Context-driven stream routing

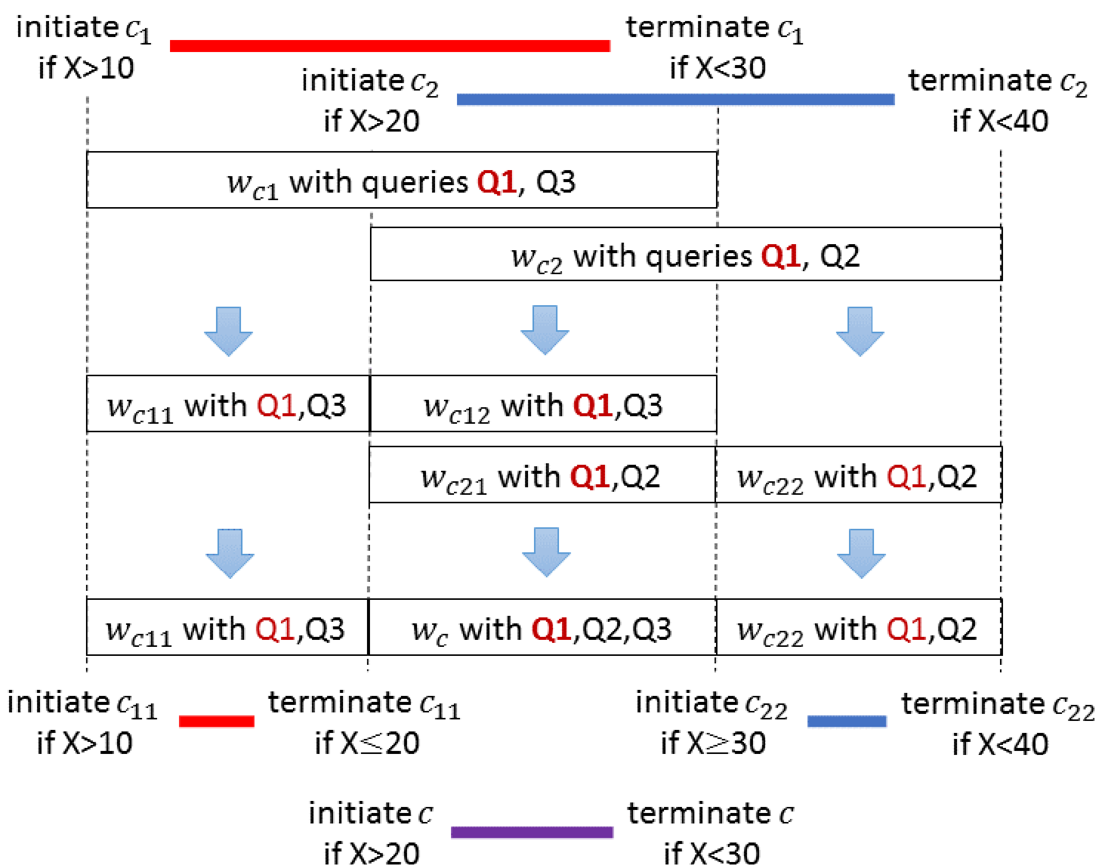
Context Workload Sharing Algorithm



Context Workload Sharing Algorithm



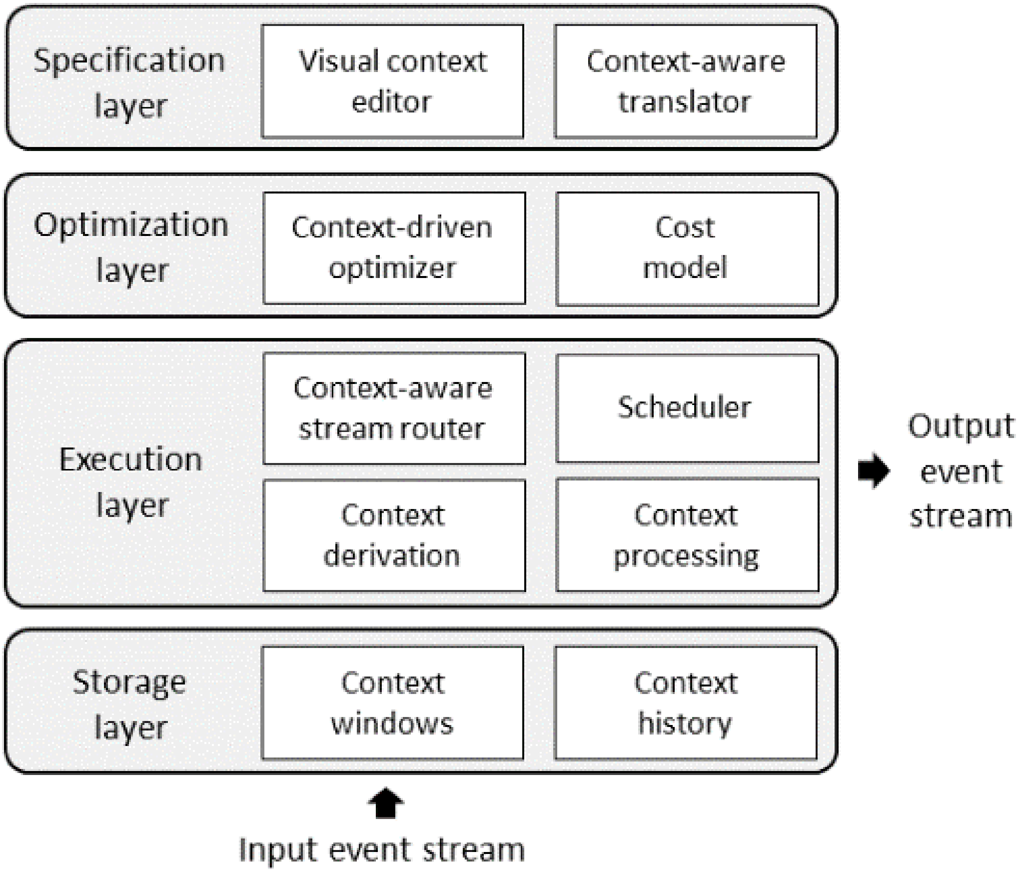
Context Workload Sharing Algorithm



Outline

CAESAR Infrastructure
& Experiments

CAESAR Architecture



Experimental Setup

Execution infrastructure:

Java 7, 1 Linux machine with 16-core

3.4 GHz CPU and 48GB of RAM

Data sets:

- Linear Road stream benchmark (LR) [1]

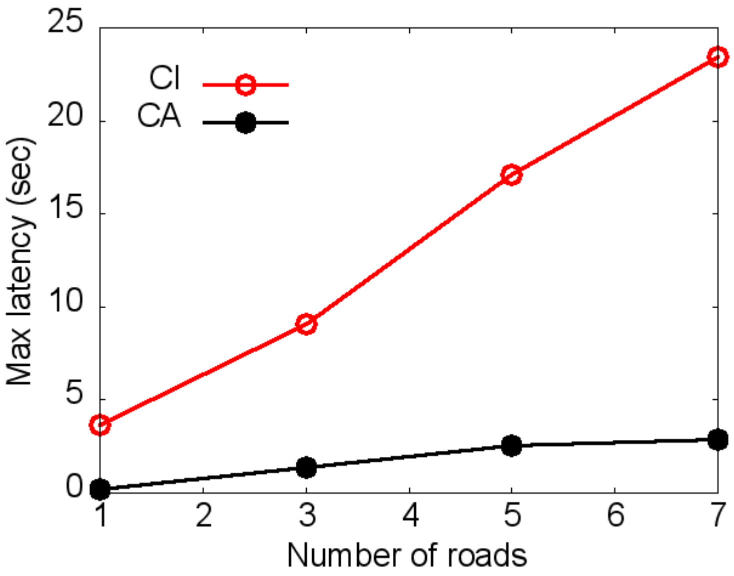
  3 roads=1.7GB

- Physical Activity Monitoring real data set (PAM) [2]

  1.6GB

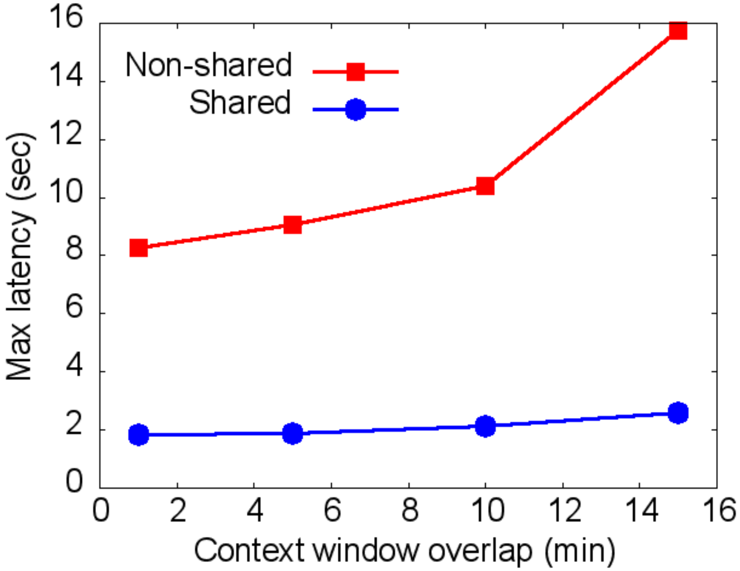
[1] A.Arasu et al., Linear Road: A stream data management benchmark. VLDB'04
[2] A.Reiss et al., Creating and benchmarking a new data set for physical activity monitoring. PETRA'12

Context-aware Event Stream Analytics



For 7 roads, context-aware (CA) event stream analytics is **9**-fold faster than context-independent (CI) approach.

Context-aware Event Query Sharing



If 30 context windows of length 15 minutes process 4 event queries each and overlap by 15 minutes, workload sharing wins **6**-fold.

Outline

Conclusions

Conclusions

- CAESAR is first context-aware CEP system
- Graphical context-specification model
- Context-aware algebra
- Context-driven optimization techniques
- Execution infrastructure
- 8-fold speed up on average

Acknowledgement

- Advisors: Elke A. Rundensteiner, Dan Dougherty
- Collaborator: Chuan Lei
- DSRG group at WPI
- EDBT reviewers
- NSF grants IIS 1018443 and IIS 1343620