IVANNIKOV MEMORIAL WORKSHOP

Towards Scalable Complex Event Processing

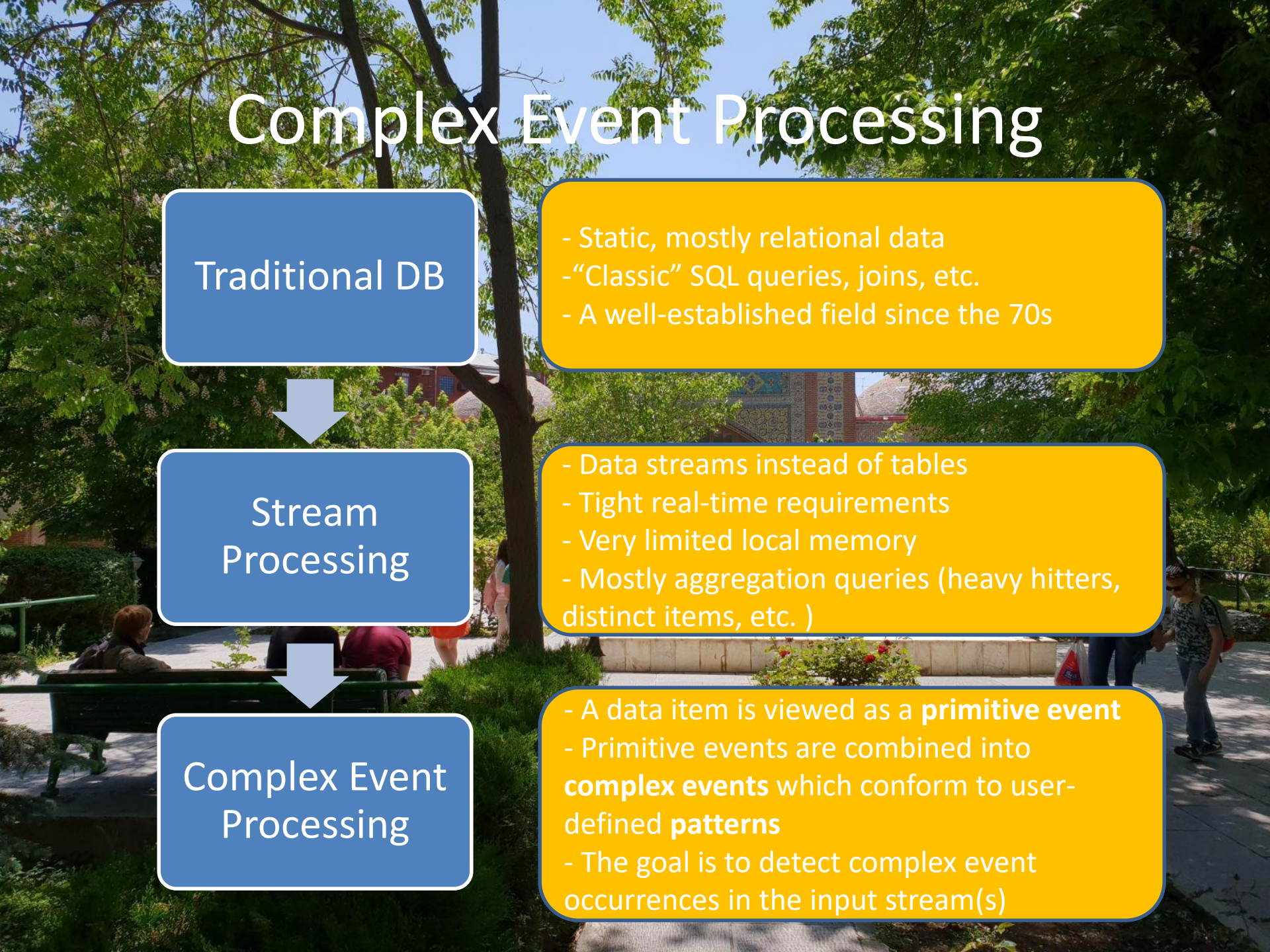
Fighting the Exponentiality of Event Pattern Detection

Ilya Kolchinsky
Assaf Schuster

Overview

- Complex event processing (CEP)
- The reason CEP is difficult
- Data-aware CEP
- Lazy evaluation in CEP
- Join methods for CEP
- Adaptive CEP

Complex Event Processing



CEP Example 1 – Security Surveillance System

SEQ(MainLobbyCameraEvent a,
CorridorCameraEvent b,
RestrictedAreaCameraEvent c)
WHERE (a.person_id == b.person_id == c.person_id)

CEP Example 2 – Monitoring Stock Prices

SEQ(GoogleStockPriceUpdate a,

MicrosoftStockPriceUpdate b,

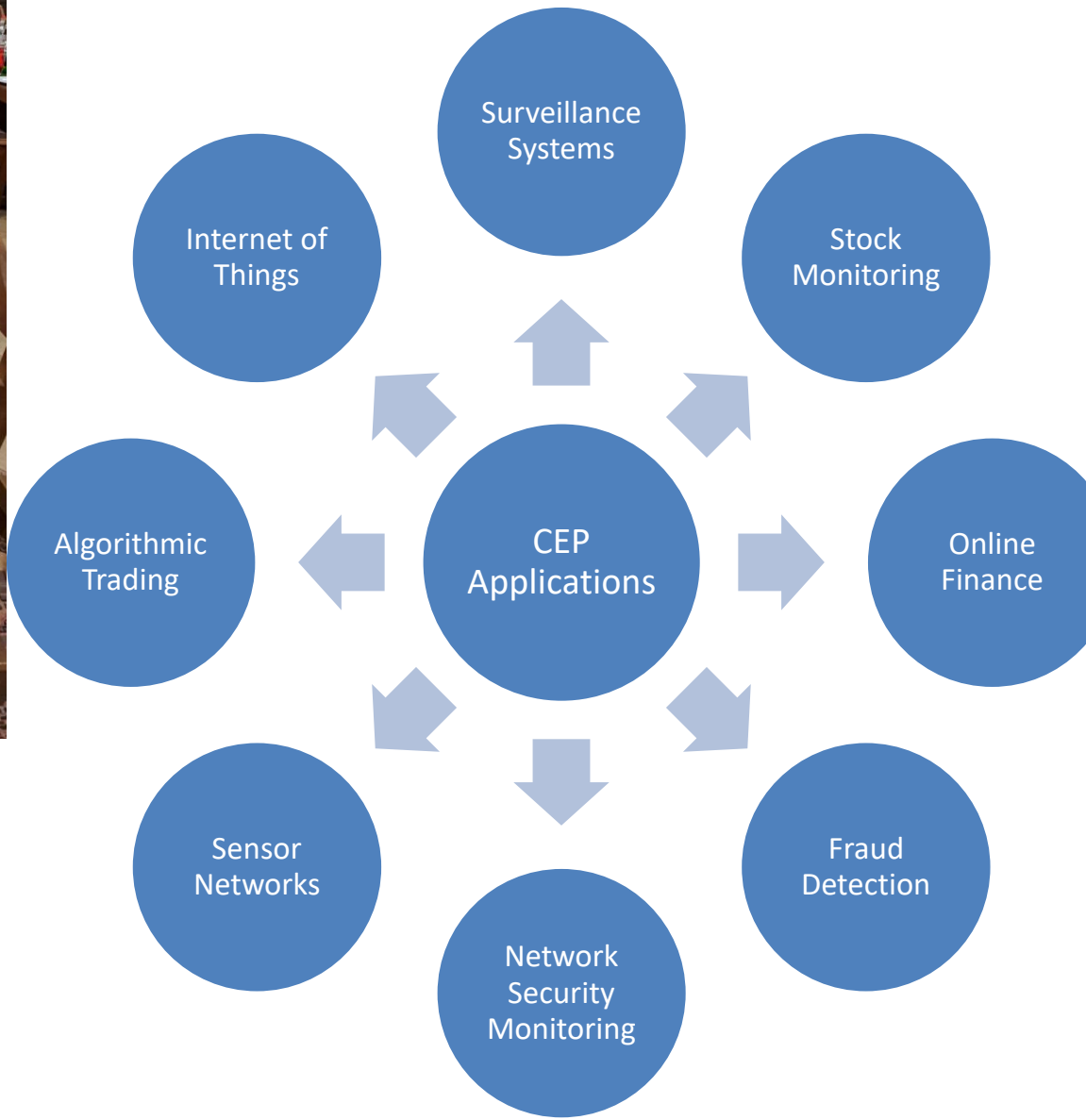
AppleStockPriceUpdate c)

WHERE

((a.price < b.price) AND (b.price < c.price))

WITHIN 10 minutes

CEP Applications



Pattern Types in CEP

Disjunctions

- at least one of the specified events must be occur

Conjunctions

- all events specified in the pattern must occur

Sequences

- all events must occur in the predefined order

Pattern Types in CEP – contd.

Negations

- some events are prohibited from occurring at the specified positions

Kleene closure

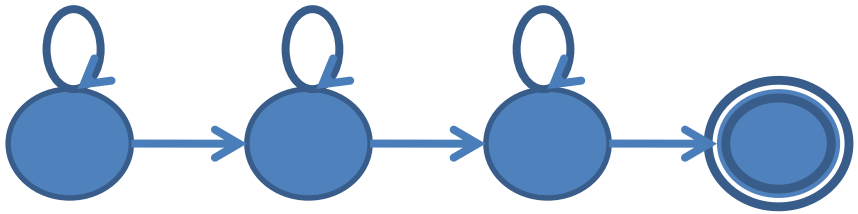
- some events may appear an unlimited number of times

Nested patterns

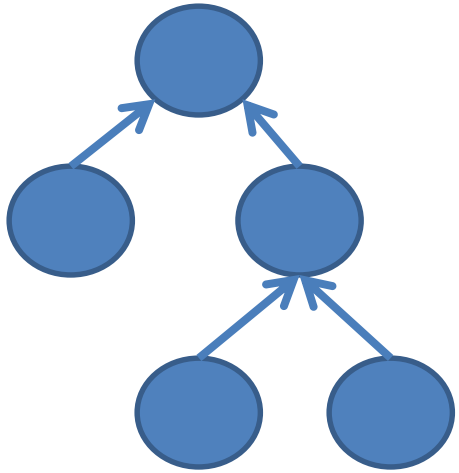
- arbitrary combinations of all of the above

CEP Evaluation Mechanisms

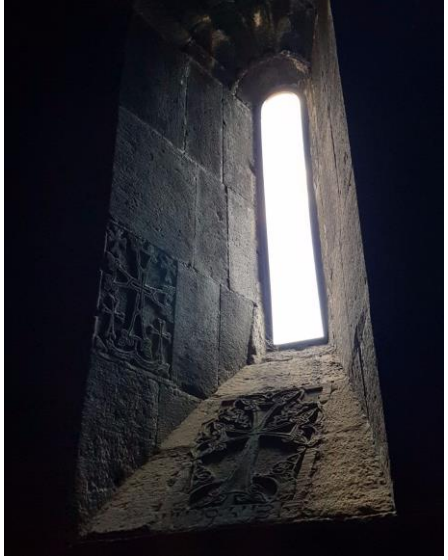
- Non-deterministic Finite Automata

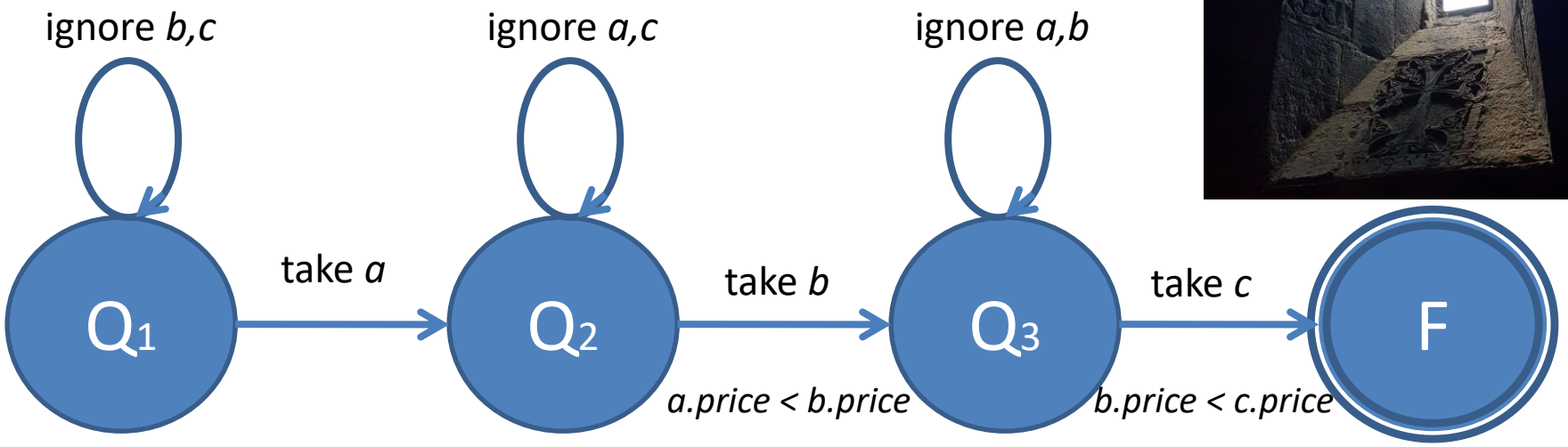


- Evaluation Trees



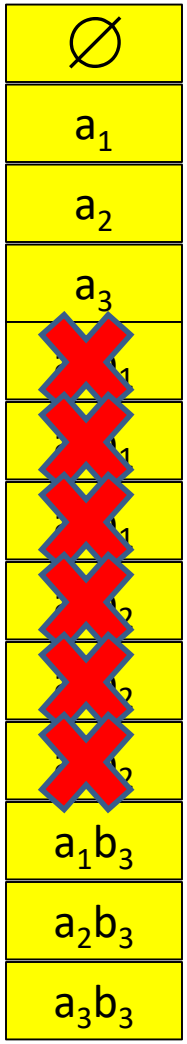
NFA for Pattern From Example 2



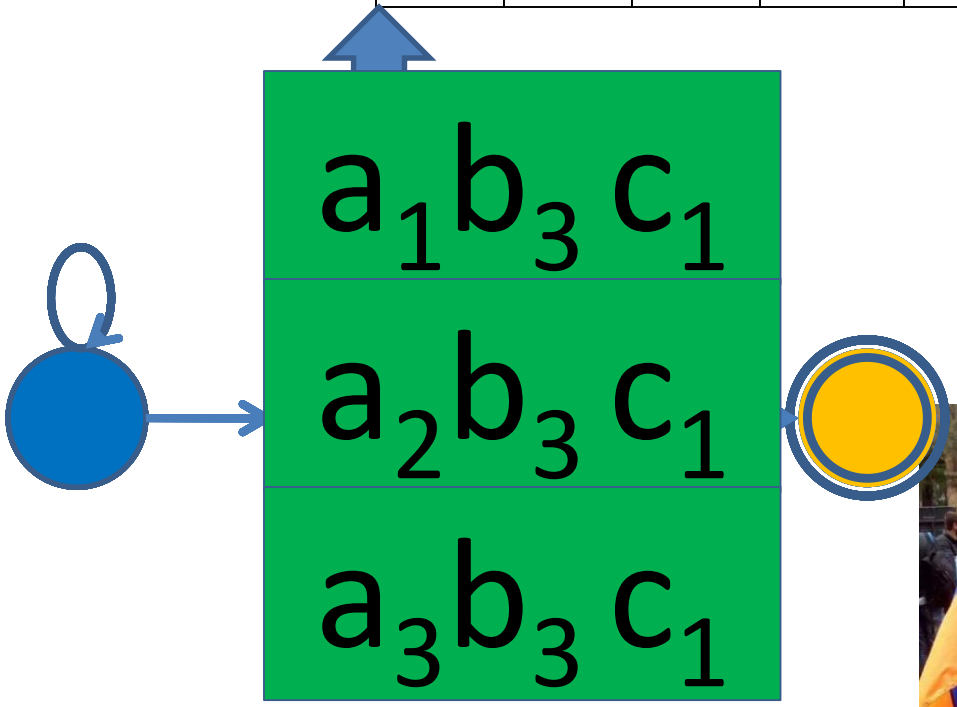


SEQ(GoogleStockPriceUpdate a, MicrosoftStockPriceUpdate b,
AppleStockPriceUpdate c)
WHERE ((a.price < b.price) AND (b.price < c.price))
WITHIN 10 minutes

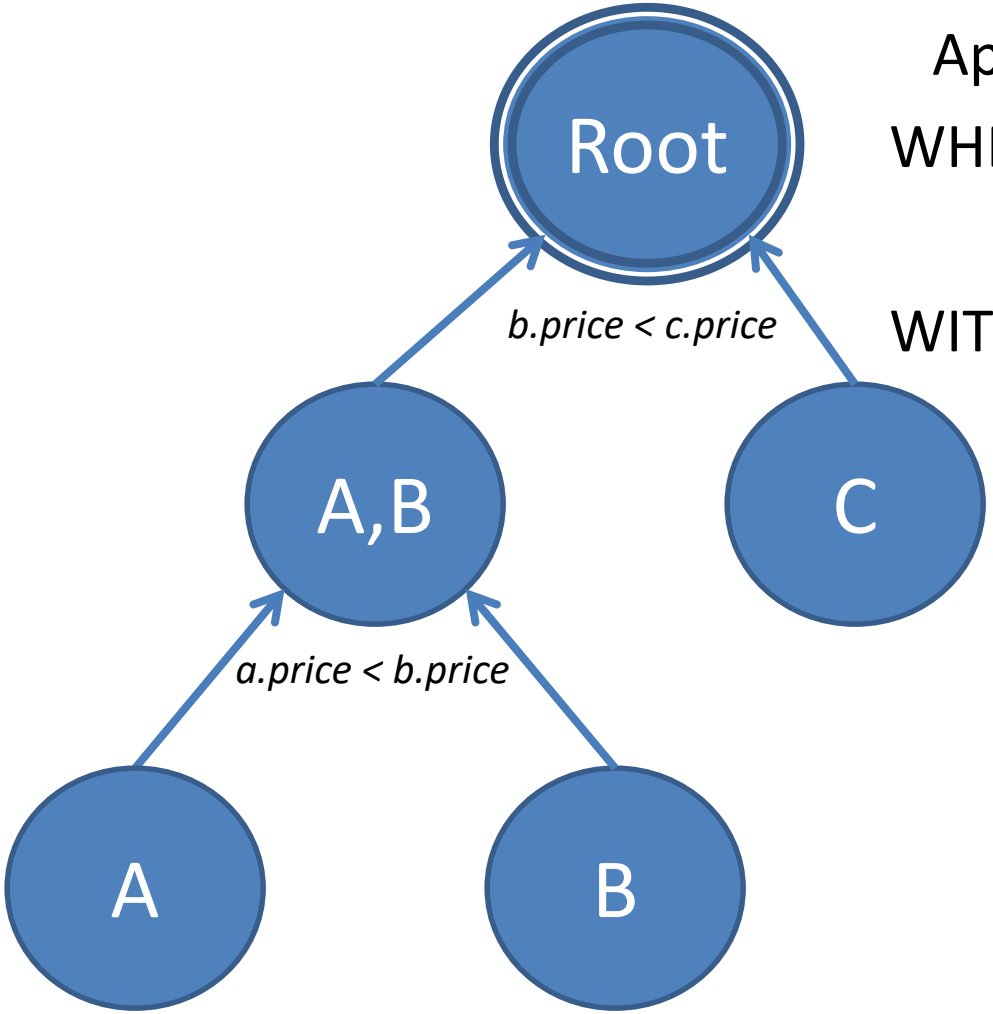
NFA-Based Detection Example



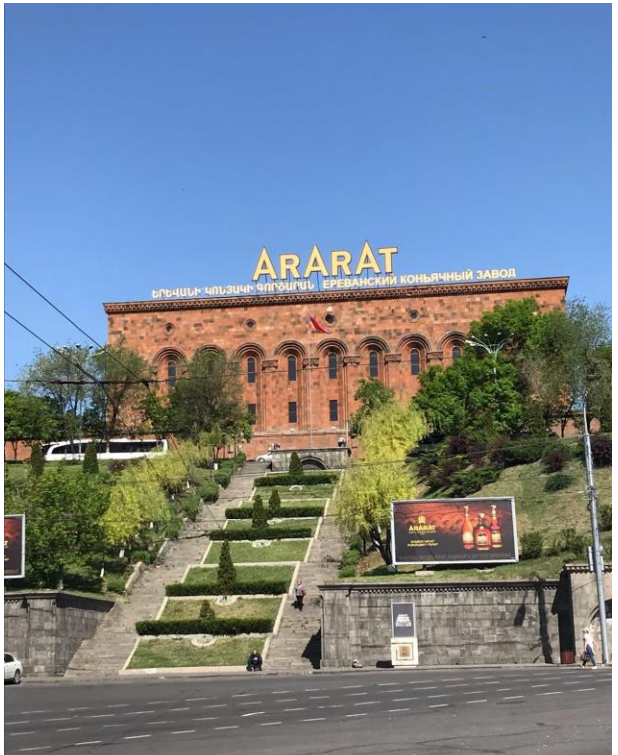
$a_1^{p=1}$	$a_2^{p=1}$	$a_3^{p=4}$	$b_1^{p=7}$	$b_2^{p=9}$	$b_3^{p=5}$	$c_1^{p=6}$



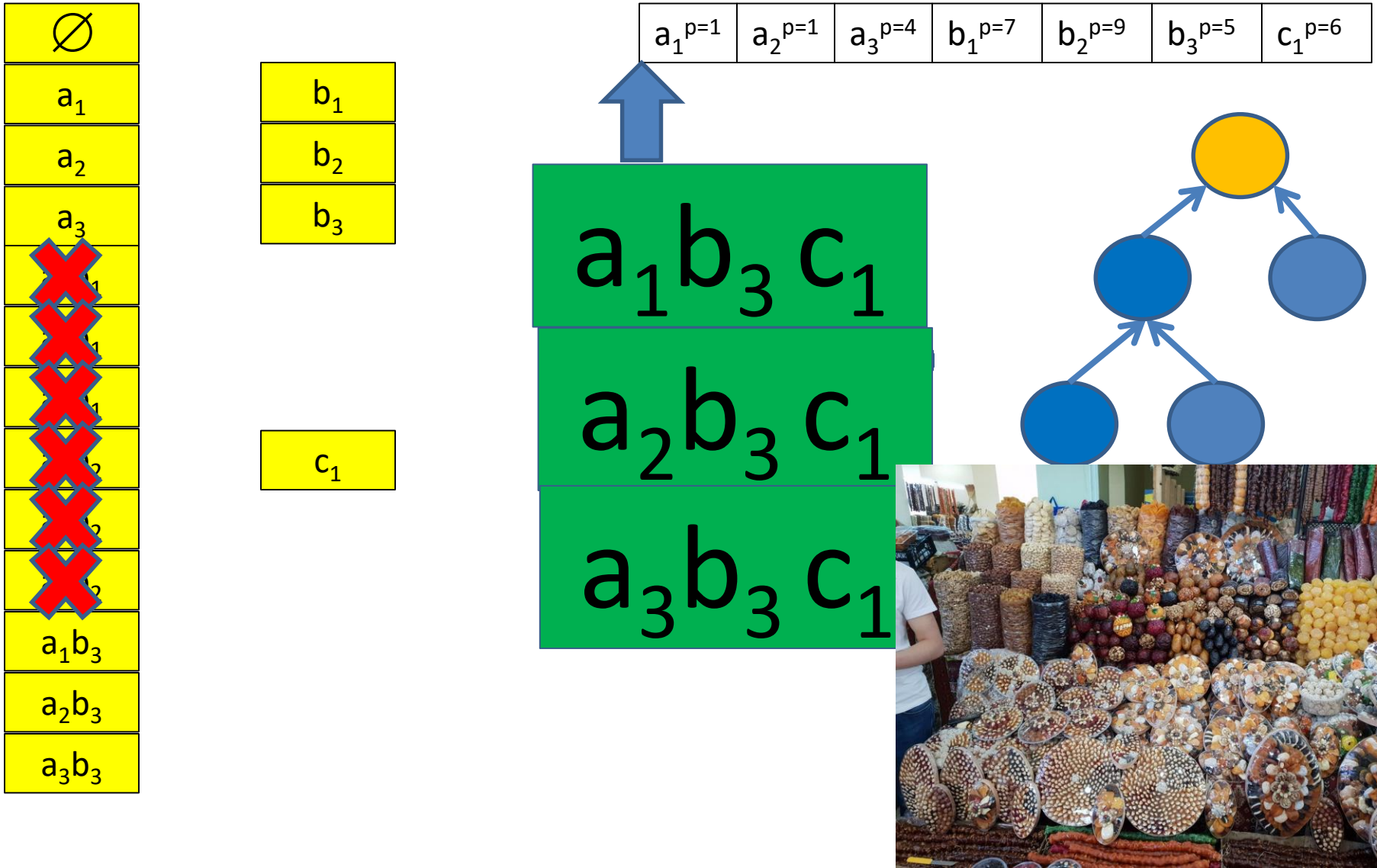
Tree Based Evaluation




```
SEQ(
   GoogleStockPriceUpdate a,
   MicrosoftStockPriceUpdate b,
   AppleStockPriceUpdate c)
WHERE ((a.price < b.price) AND
          (b.price < c.price))
WITHIN 10 minutes
```

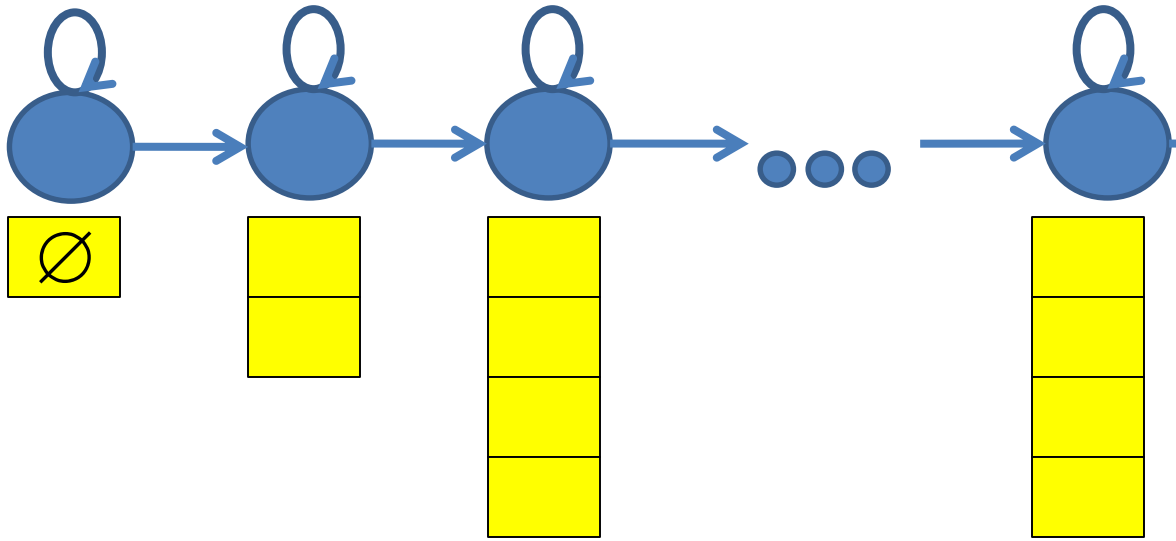


Tree-Based Detection Example



Complex Event Processing is difficult

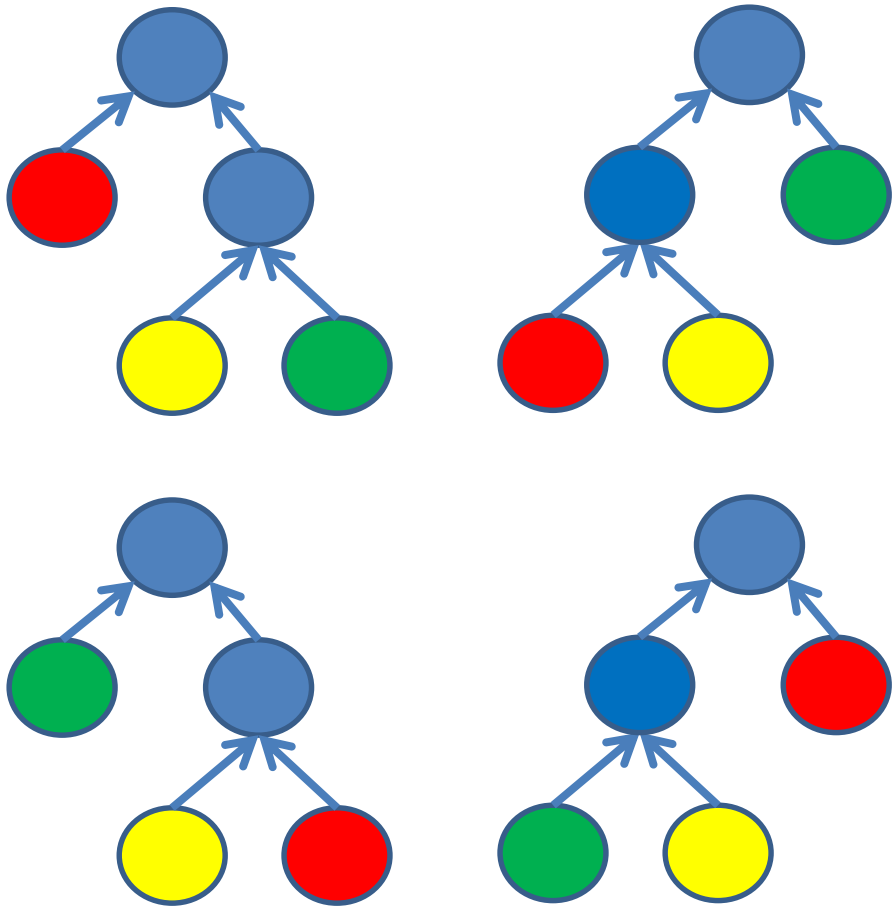
- The number of partial matches to be maintained during the detection process is **exponential in pattern size!**



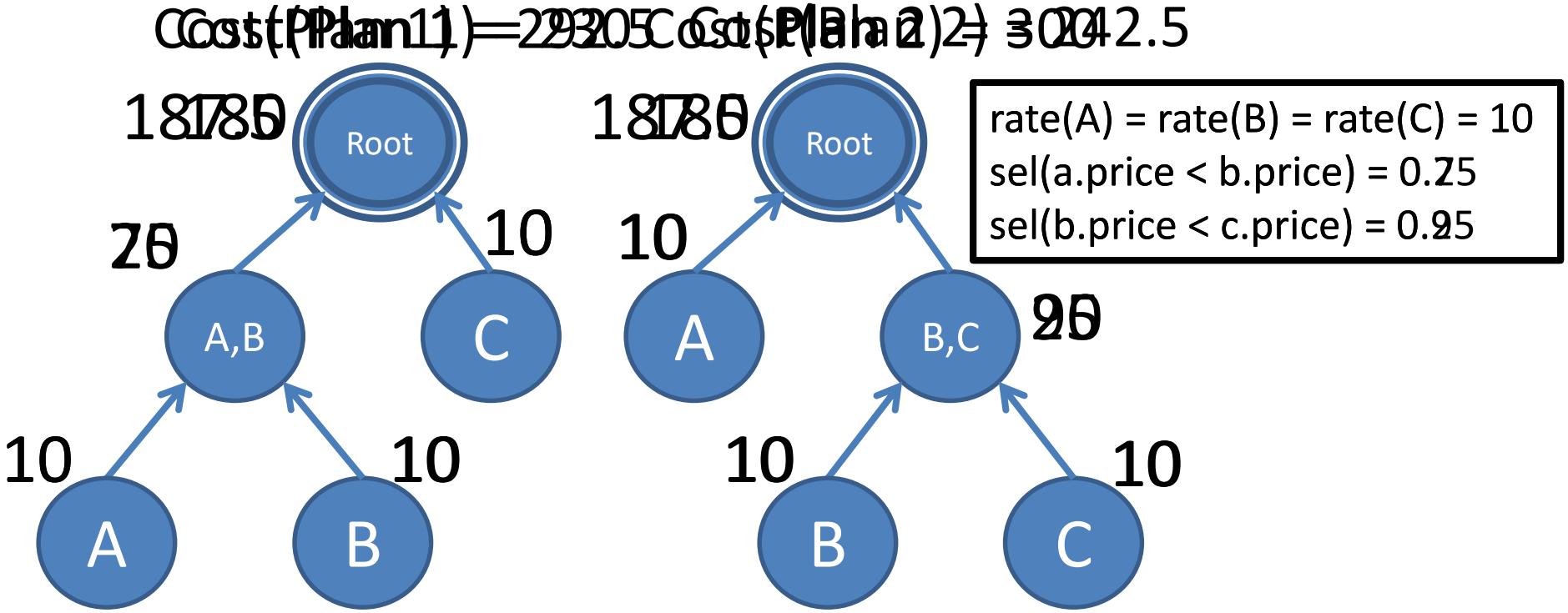
Traditional human-generated patterns are simple and short…
But recent machine-generated patterns
are more sophisticated – an algorithmic challenge!

Data-Aware CEP

- Define multiple evaluation plans for a given pattern
- Leverage the available knowledge on statistical data properties to select the best plan
- Possible for trees ☺
- No such model for NFA ☹



Data-Aware CEP Example



SEQ(A a, B b, C c)

WHERE (a.price < b.price) AND (b.price < c.price)

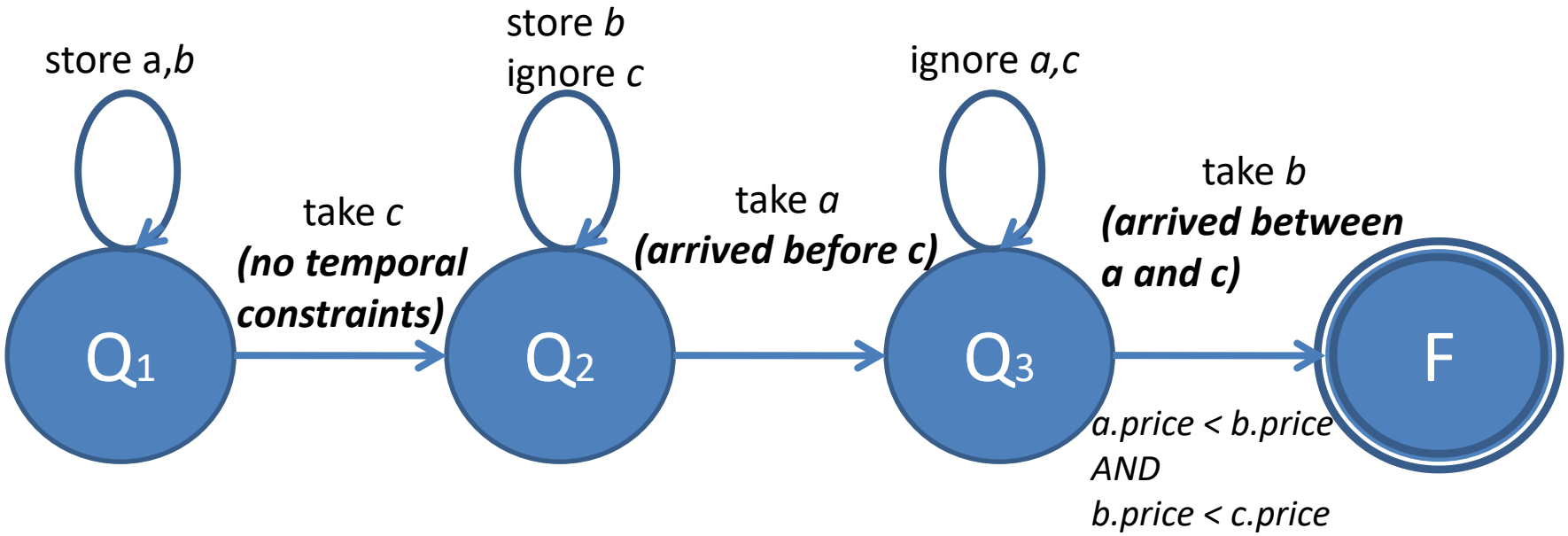
WITHIN 1 hour

Lazy Evaluation Model for NFA

[DEBS'15 Best Paper Award]

- Process incoming events according to arbitrary order (rather than in order of appearance)
- Keep the unprocessed events in an intermediate storage
- Process the buffered events only when required by the plan in use

Lazy NFA – (order *c,a,b*)

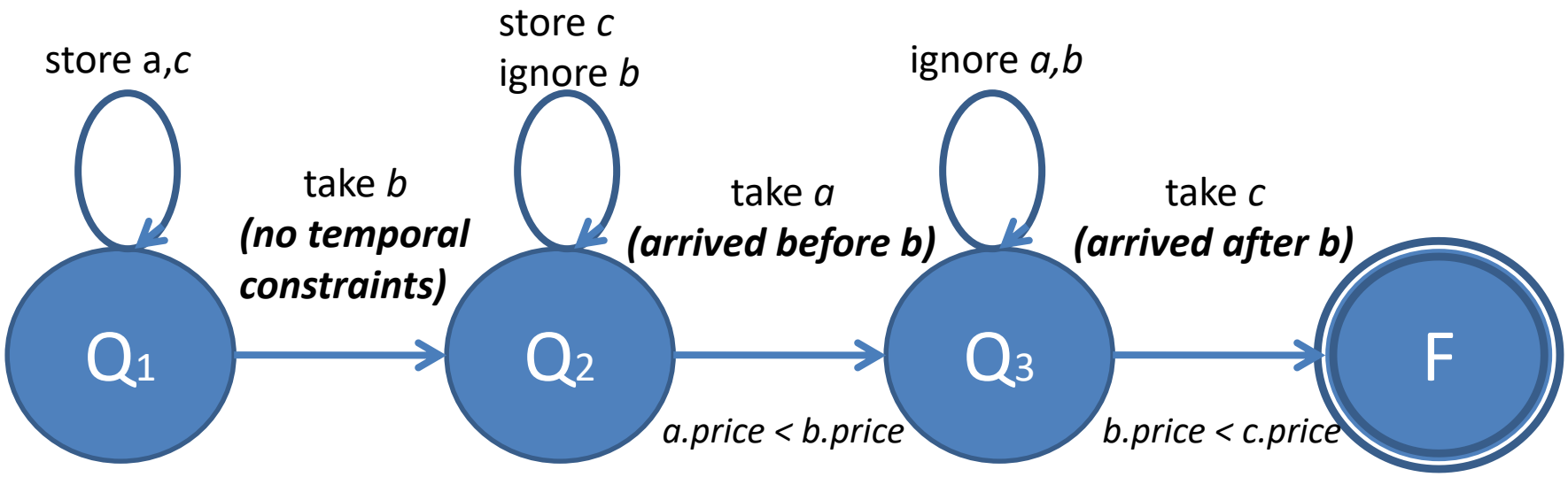


SEQ(A a, B b, C c)

WHERE (a.price < b.price) AND (b.price < c.price)

WITHIN 1 hour

Lazy NFA –(order *b,a,c*)



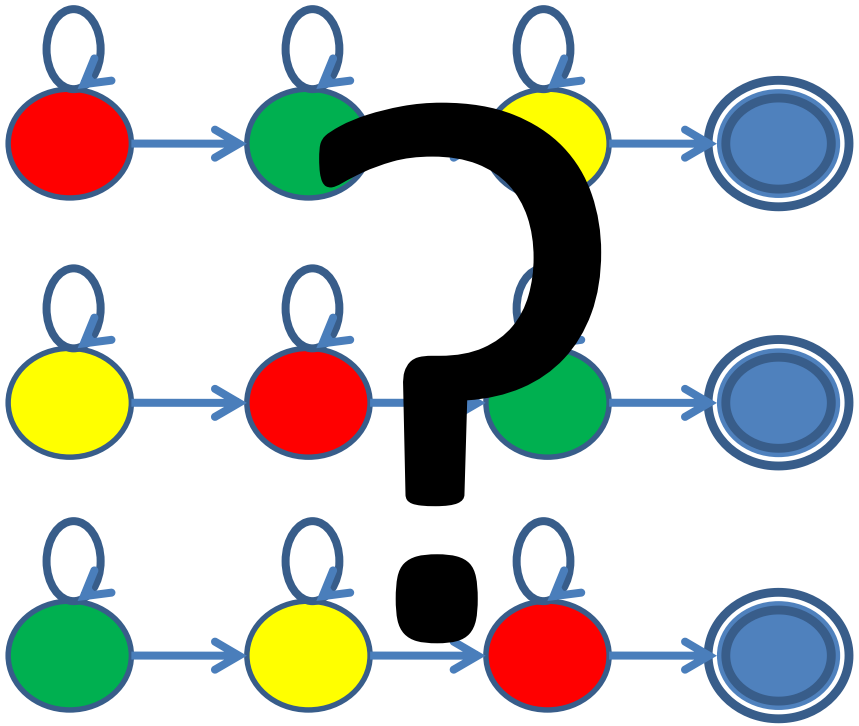
SEQ(A a, B b, C c)

WHERE (a.price < b.price) AND (b.price < c.price)

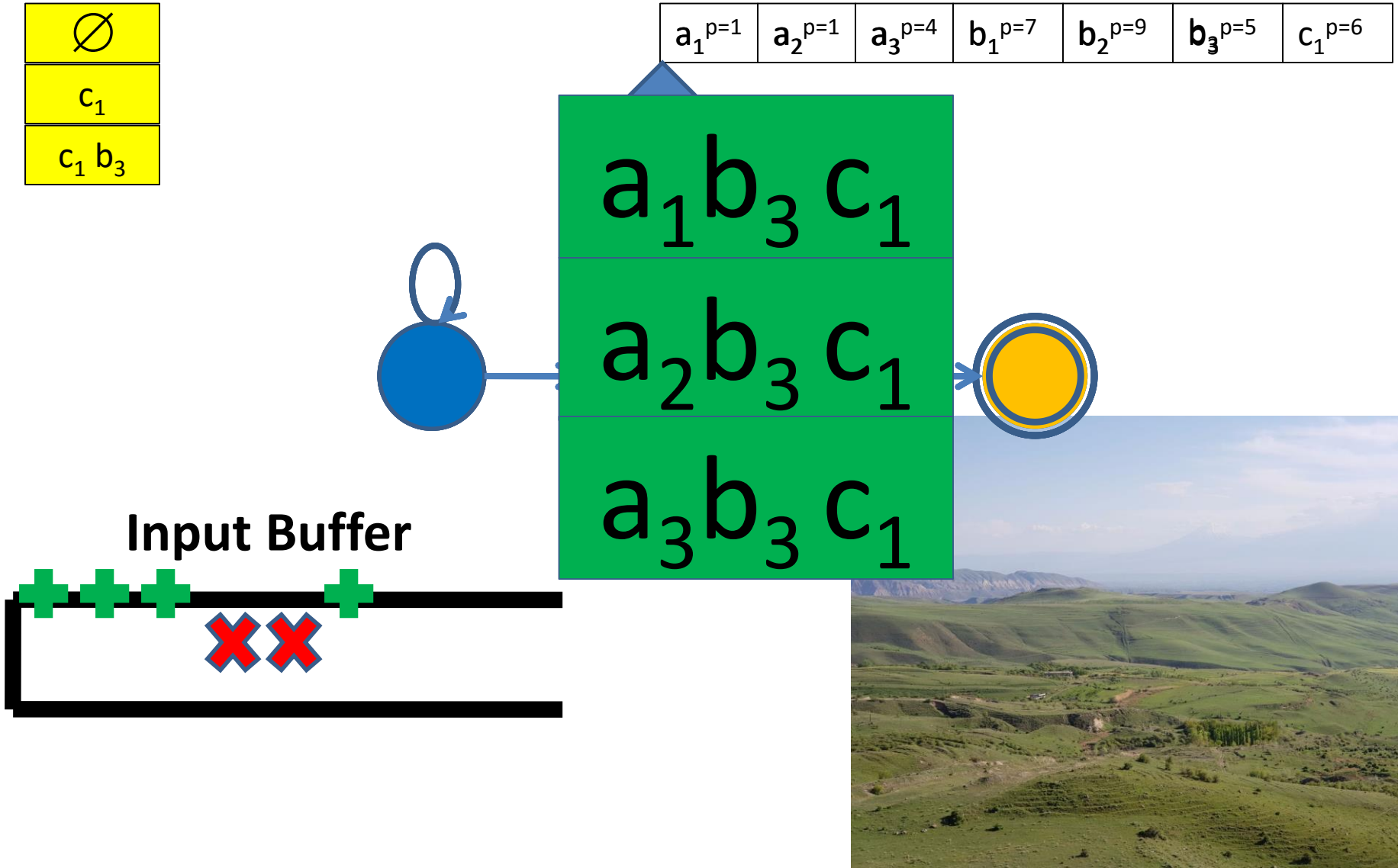
WITHIN 1 hour

What is the best evaluation order?

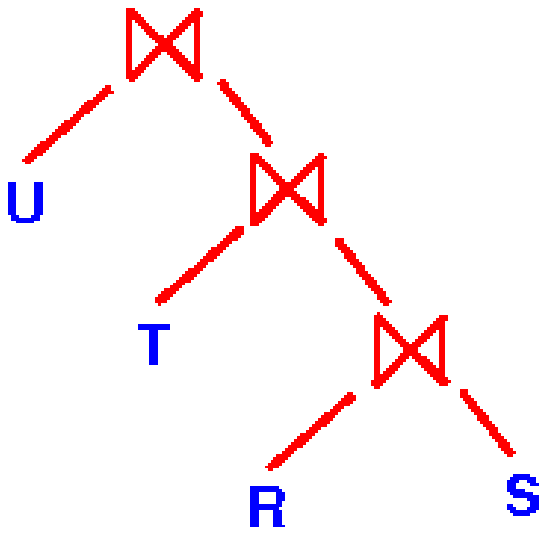
- A simple solution: order the events from the rarest to the most frequent
  - assume the frequencies are given and are not changing



Lazy NFA Pattern Detection Example (order *c,b,a*)



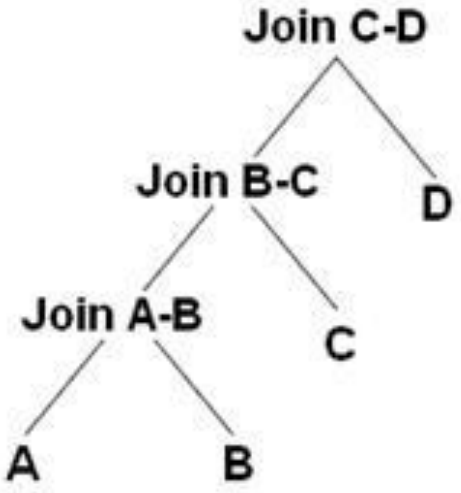
Join Order Estimation



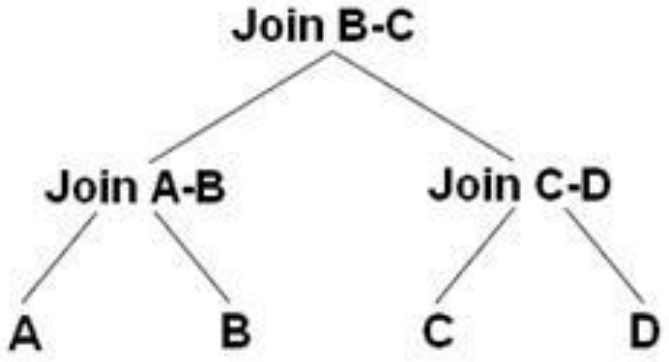
- A well-known problem since the 80's
- Given *n* relations joining on *m* attributes, what is the most efficient way to perform the join?
- Extensively studied, lots of algorithms published

- Hmmmm…. So…….
- Can an instance of CEP plan generation problem be transformed into this problem?
- [Submission-I 2018]

Join Query Plan Types

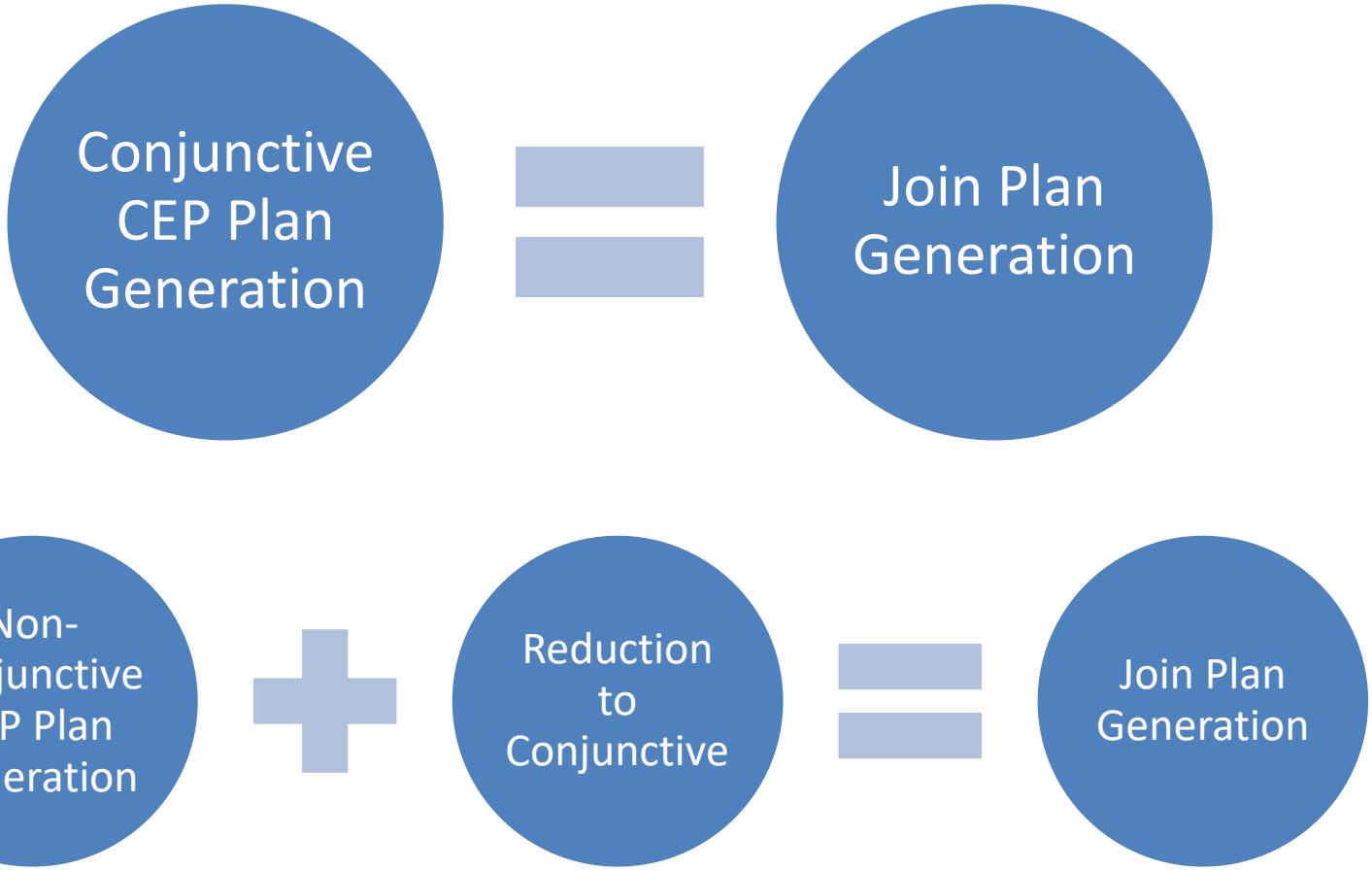


Left Deep Tree
(resembles NFA?)



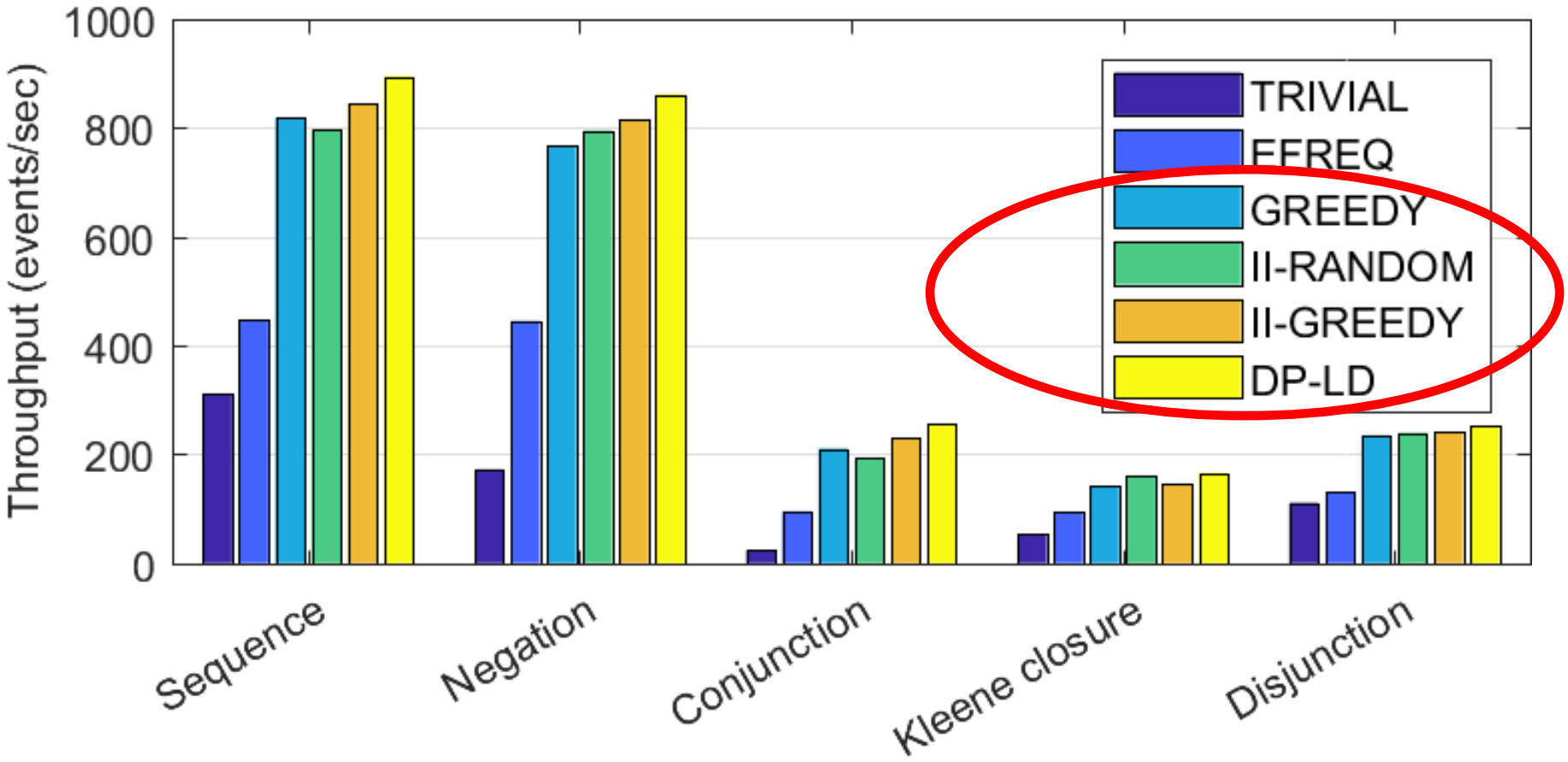
Bushy Tree
(resembles CEP Trees?)

Problem Equivalence



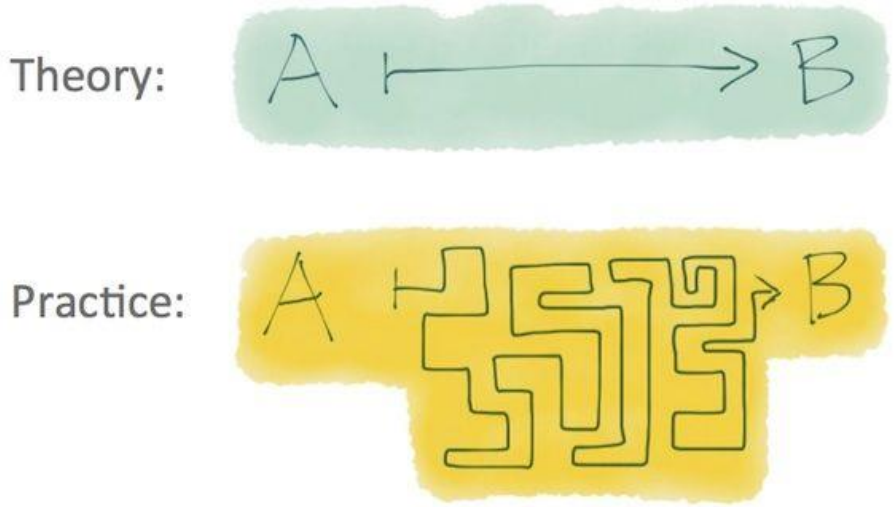
Existing Algorithms

Experimental Results – NFA throughput

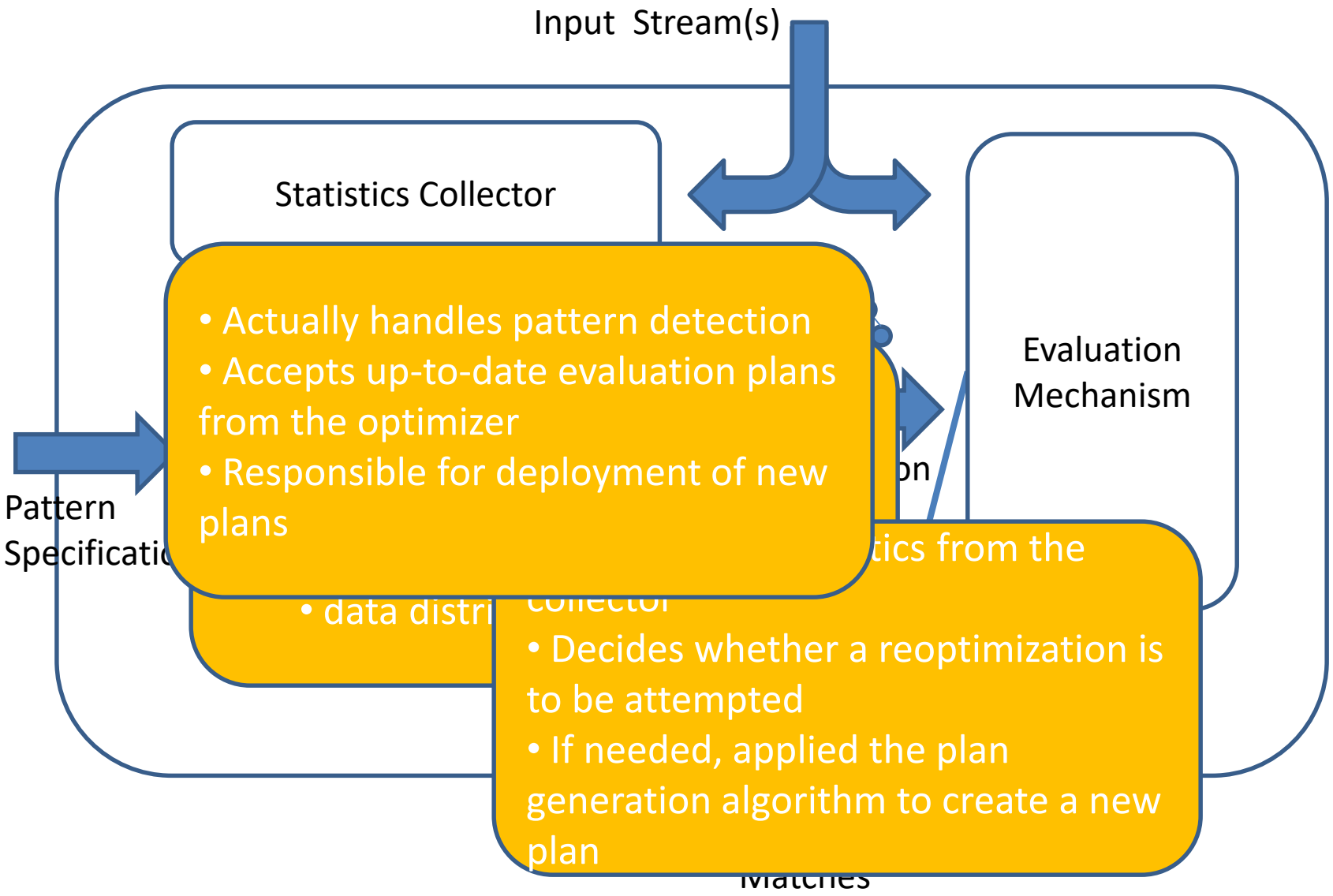


Data-Aware CEP in Practice

- Will the data-aware method work in a real-life scenario?
- Probably not
  - arrival rates / selectivities not known in advance
  - arrival rates / selectivities subject to frequent changes
- **Adaptation** is needed



Adaptive CEP



Possible Adaptation Strategies

- Periodically recalculate the evaluation plan based on new statistics
- Monitor all statistic values against a predefined threshold

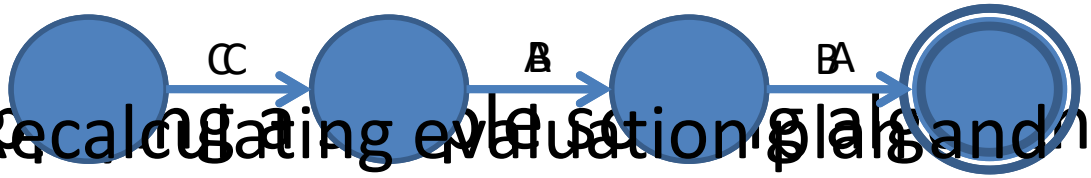
- Can we do better?

Invariant-based Method [Submission-II 2018]

SEQ(A a, B b, C c)

WHERE (a.price < b.price) AND (b.price < c.price)

WITHIN 1 hour



Arriving events trigger recalculating evaluation plans and checking invariants…

Is rate(B) > rate(A)?

Invariants
~~rate(B)>rate(C)~~
~~rate(B)>rate(A)~~

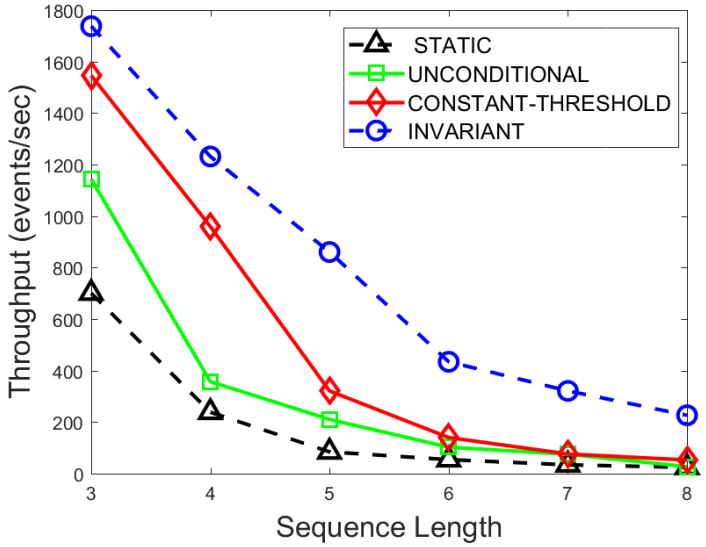
Tighter condition found!

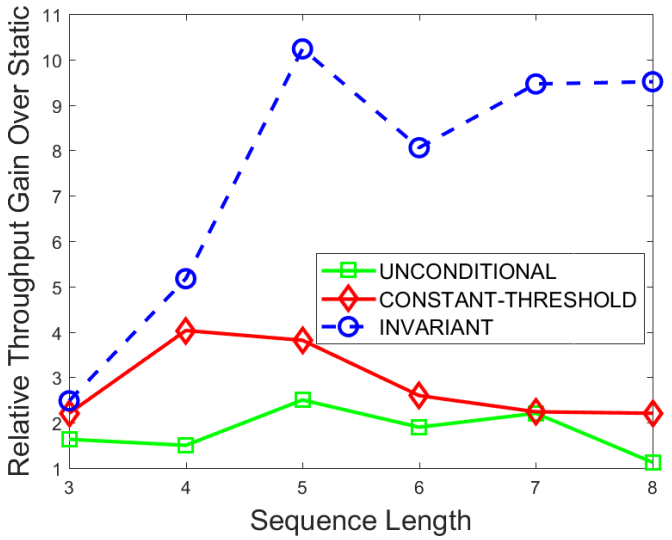
rate(A) = 20

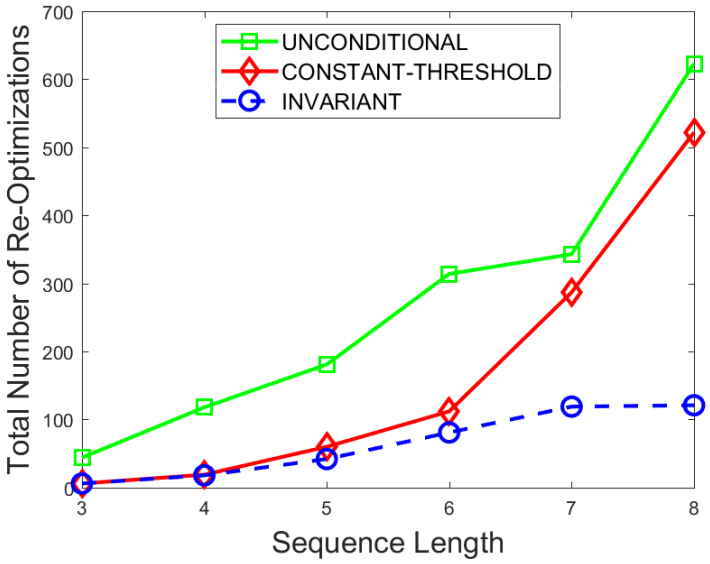
rate(B) = 50

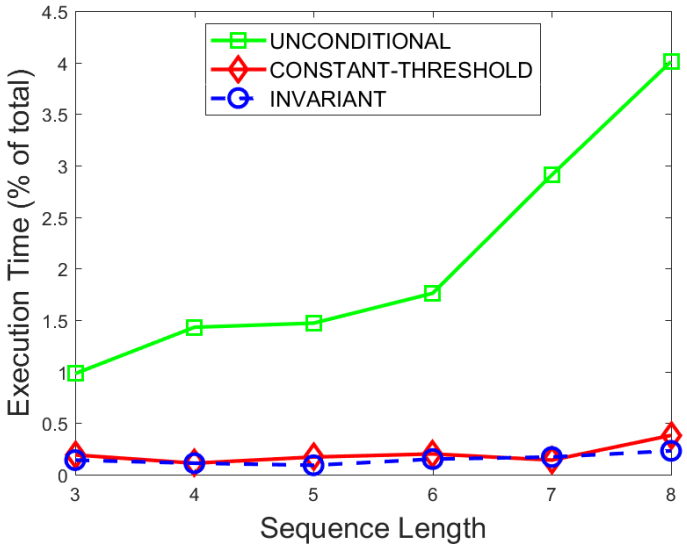
rate(C) = 10

Experimental Results - NFA





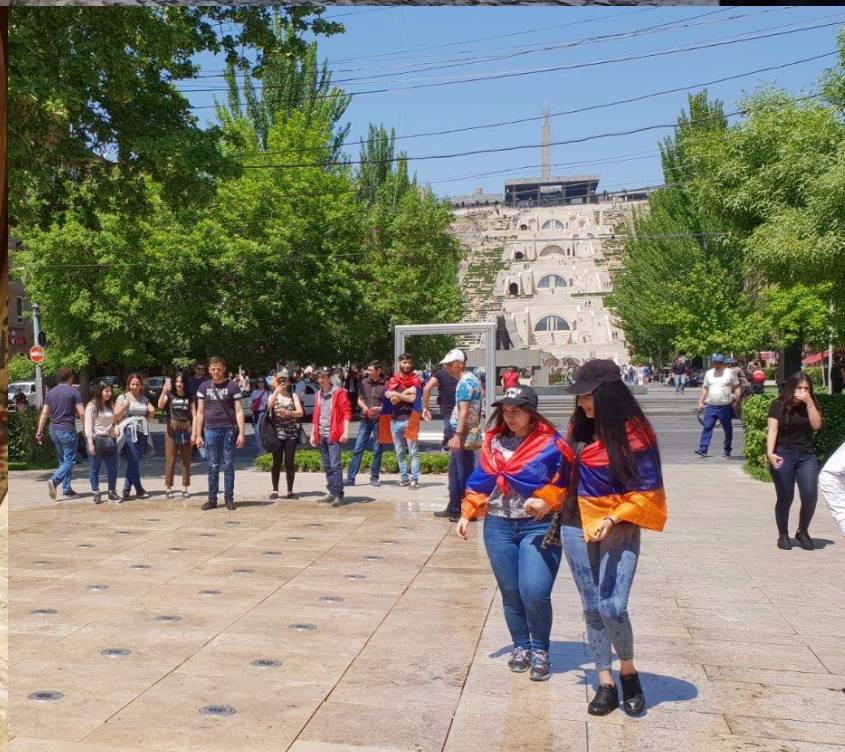




Summary

- Incorporating data statistics allows a CEP system to achieve a performance boost of orders of magnitude (Still not enough ☹)
- Detection latency and memory consumption are also significantly improved
- Many open questions remain


ARARAT
PEACE BARREL
Set for ageing in honor of the
visit of co-chairs of OSCE Minsk Group
and
will be opened when
Karabakh conflict
is resolved.
Yerevan
2001



QUESTIONS?