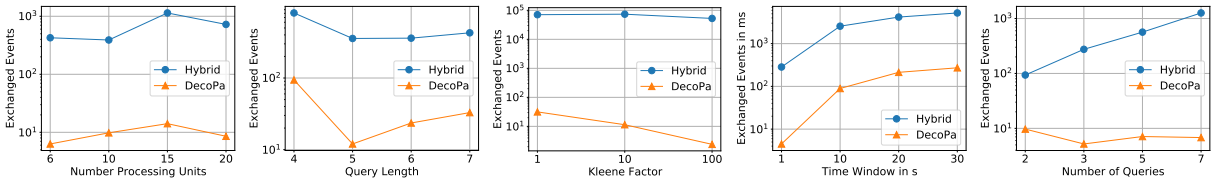
Material in Response to Comments on: "DecoPa: Query Decomposition for Parallel Complex Event Processing"

November 30, 2023

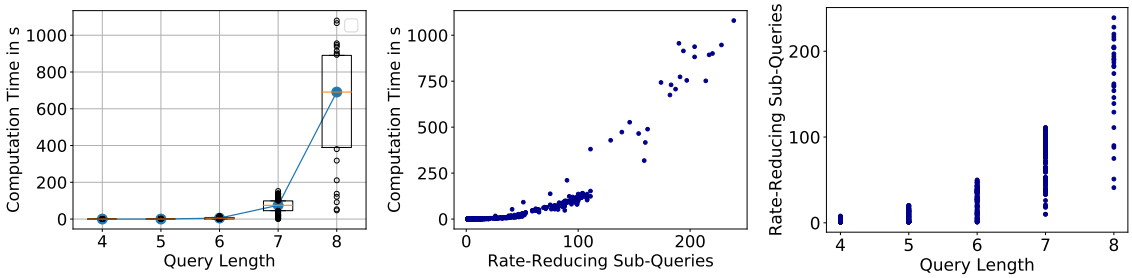
1 Results for Network/Shuffle Costs



(a) (b) (c) (d) (e)

Figure 1: Exchanged events between processing units for DecoPa and Hybrid for maximal scaling factor of Hybrid approach.

2 Results for Optimization Time



(a) (b) (c)

Figure 2: Computation time in seconds in (a) relation to query length and (b) number of sub-queries considered for decomposition. (c) Number of considered sub-queries per query length.

3 Query Snippets of Real-World Data Sets

Citi Bike Types. Characterization for derivation of event types:

(1) short trip $\leq 100s$

(2) long trip $100s \leq xleq5000s$

(3) very long trip $\geq 5000s$

(4) old, year of birth ≤ 1962

(5) young, year of birth ≥ 1962

(6) customer (instead of member) if not year of birth given

Examples for Event Types: LongC corresponds to a long trip with the driver being a customer. ShortY corresponds to a short trip with the driver being a young member. VLongO corresponds to a very long trip with the driver being a member of age group old.

Google Cluster Types. A description of the event types contained in the Google Cluster queries can be found at https://drive.google.com/file/d/10r6cnJ5cJ89fPWCgj7j4LtLBqYN9RiI9.

```
Query 1 - Citi Bike:
  PATTERN AND(D e1,C 2,A 3,B 4,H 5)
  WHERE ∀(i, j) ∈, dist(i.startLoc, j.startLoc) ≥ 6km
  WITHIN 24h

Query 2 - Citi Bike:
  PATTERN AND(E , KL(D), A , H)
  WHERE ∀(i, j) ∈, dist(i.startLoc, j.startLoc) ≥ 6km
  WITHIN 24h

Query 3 - Google Cluster:
  PATTERN AND(Submit, SEQ(Evict,Enable), Finish)
  WHERE ∀(i, j)
  WITHIN 10min

  Query 4 - Google Cluster:
  PATTERN AND(KL(Schedule s), Queue l,Lost q)
  WHERE ∀(i, j) ∈ {s, l, q} × {s, l, q}, i.mem_usage ≥ j.mem_usage ∧ i.cpu_usage ≥ j.cpu_usage
  WITHIN 10min
```

Figure 3: Queries used in evaluation for Citi Bike and Google Cluster traces.