

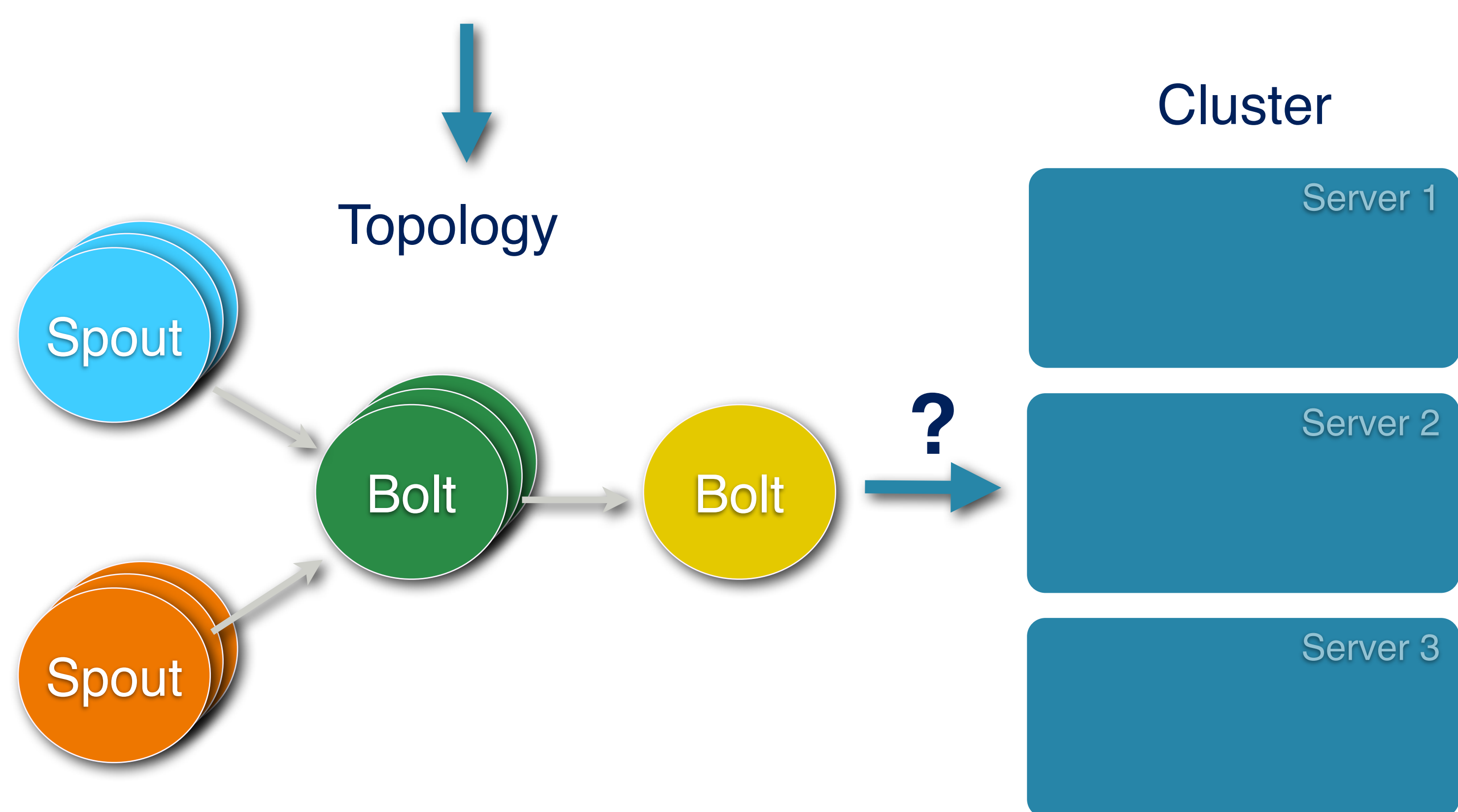


Network-Aware Workload Scheduling for Scalable Linked Data Stream Processing

Problem

Query

```
REGISTER QUERY PublicSpendingStock AS
SELECT { ?company_name ?agency_name ?contract_id ?min_price ?max_price ?factor }
FROM STREAM <wrds.crsp/ticker.trdf> [RANGE 20 DAY STEP 1 DAY]
FROM STREAM <usaspending.org/contracts.trdf> [RANGE 20 DAY STEP 1 DAY]
WHERE { GRAPH <wrds.crsp/ticker.trdf> {
  ?ticker_id wc:PRC ?ticker_price ;
  wc:COMNAM ?company_name ;
  wc:TICKER ?ticker_symbol .
} UNION GRAPH <usaspending.org/contracts.trdf> {
  ?contract_id us:agencyid ?agency_name ;
  us:obligatedamount ?contract_amount ;
  us:vendorname ?company_name .
}}
AGGREGATE { (?min_price, MIN, {?ticker_price}) }
AGGREGATE { (?max_price, MAX, {?ticker_price}) }
BIND (?max_price / ?min_price AS ?factor)
FILTER(?factor > 2)
```

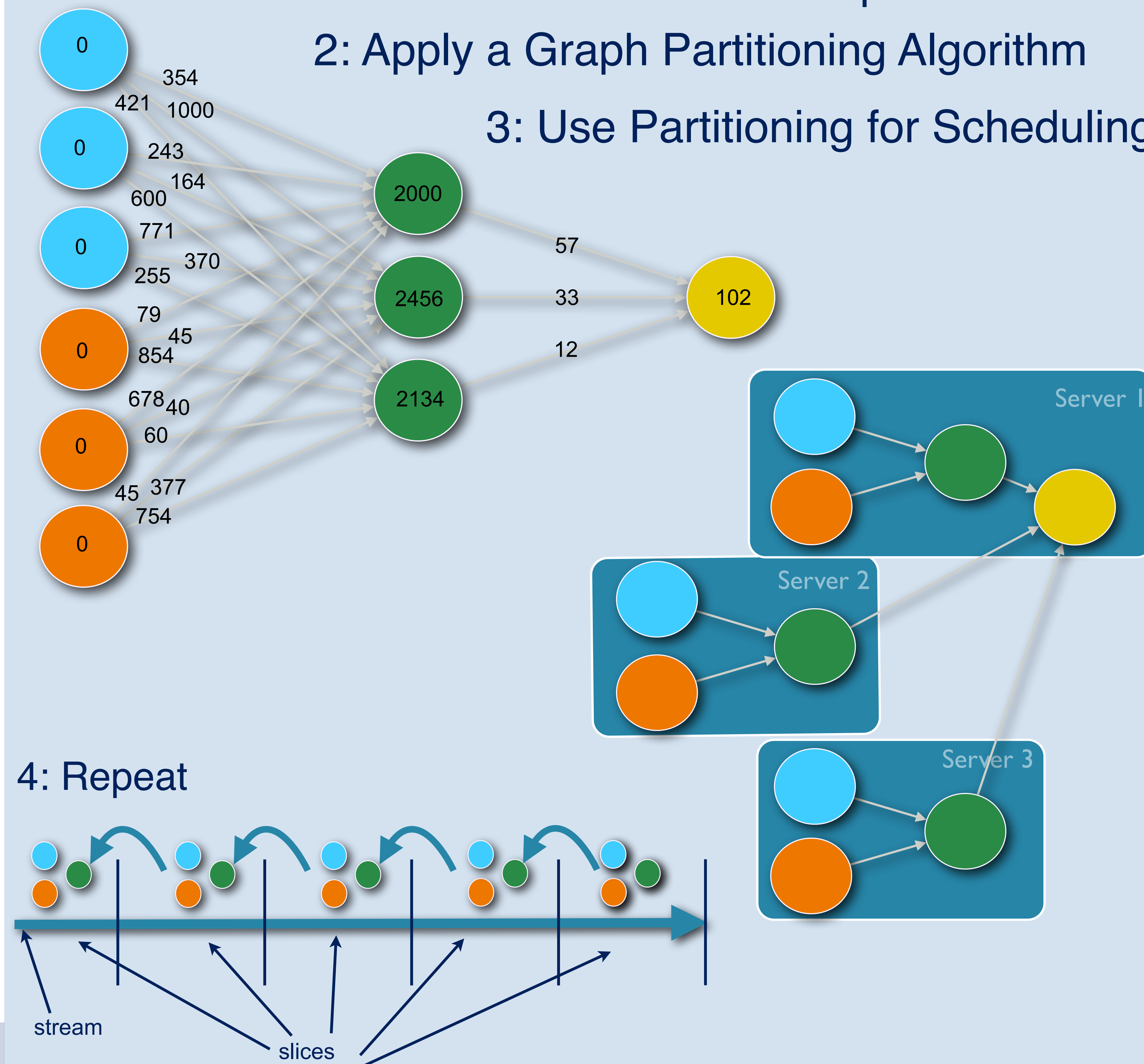


Approach

1: Record the "Communication Graph"

2: Apply a Graph Partitioning Algorithm

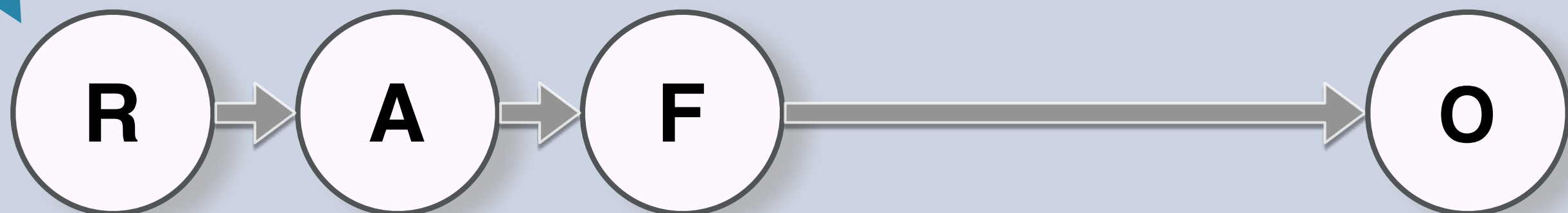
3: Use Partitioning for Scheduling



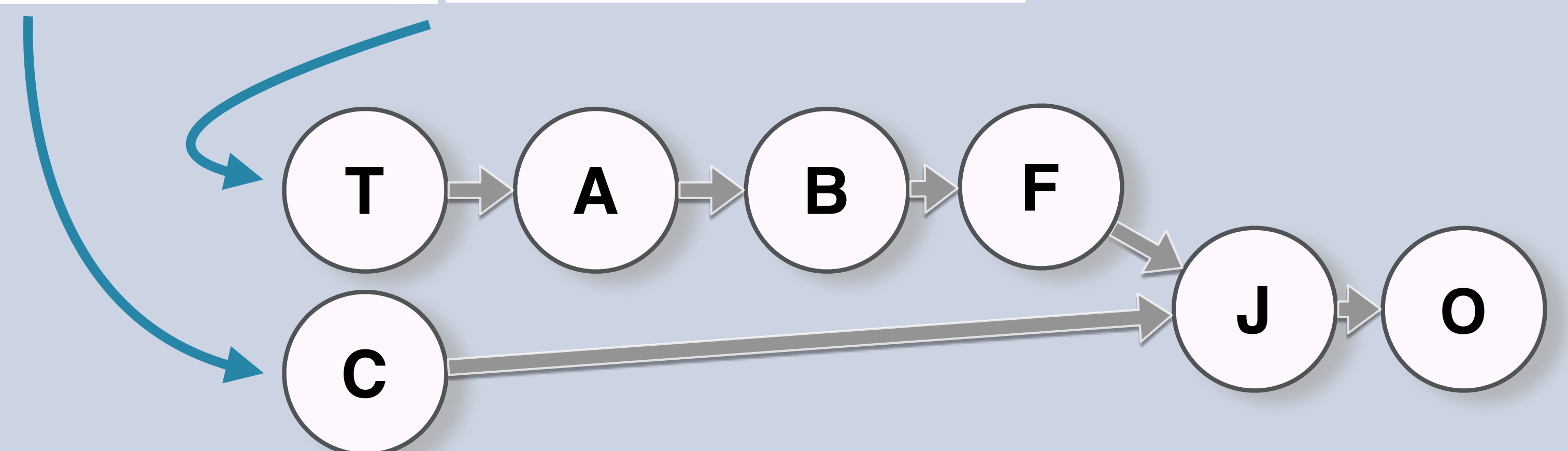
Results



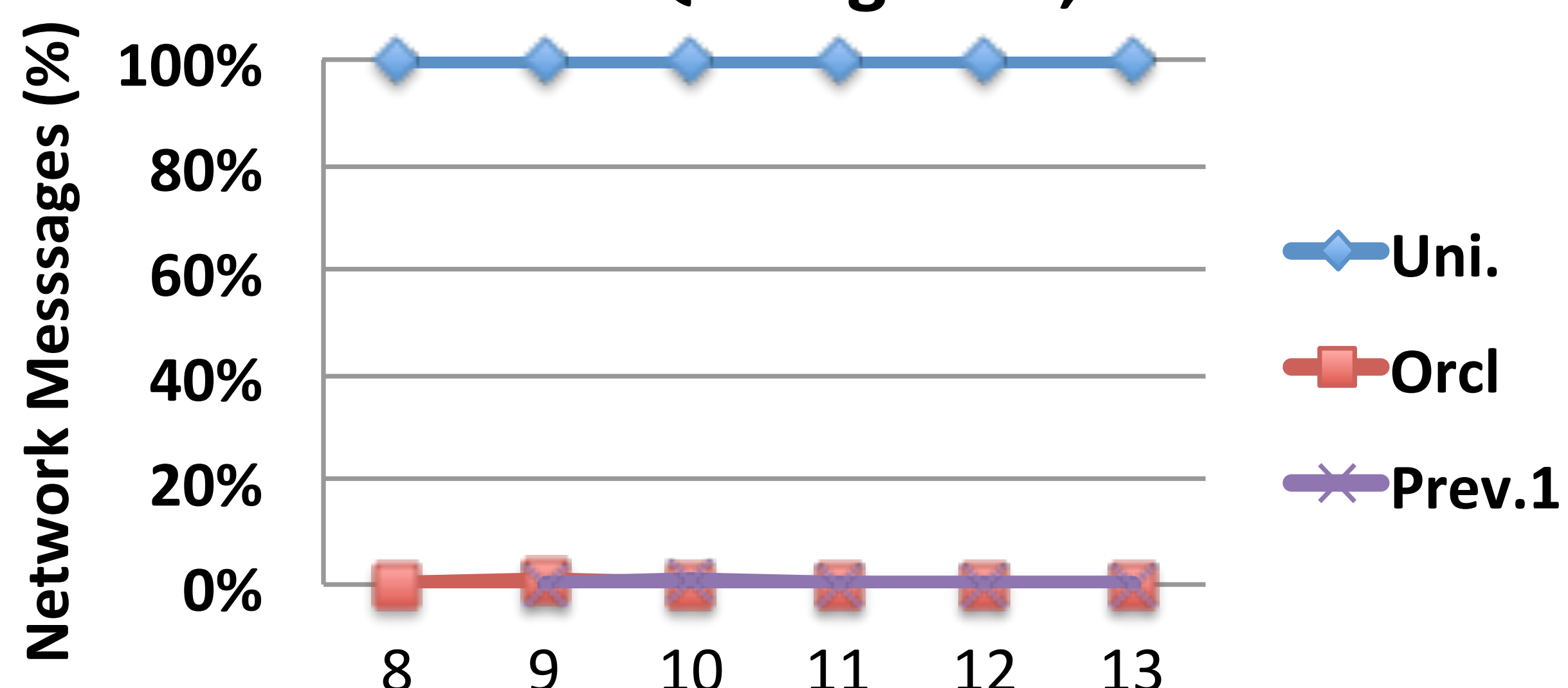
SRBench Q3



OpenGov



SRBench Q3 Aug 8-13, 2004



OpenGov 2001

