Continuous Distributed Monitoring in the Evolved Packet Core

Industry Experience Report

Romaric Duvignau [1] Marina Papatriantafilou [1]
Konstantinos Peratinos [3] Eric Nordström [2] Patrik Nyman [2]

DEBS 2019, Darmstadt (June 26).

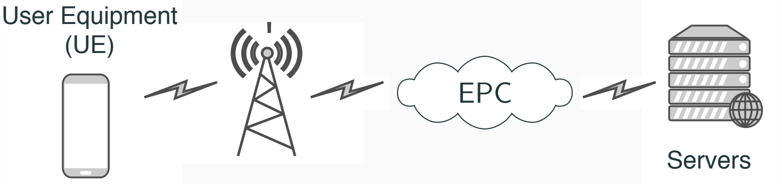
[1] Chalmers University of Technology, [2] Ericsson, [3] Chalmers student and Ericsson intern.

CHALMERS

ERICSSON

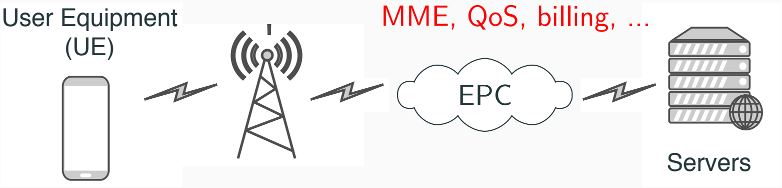
Introduction

The Evolved Packet Core



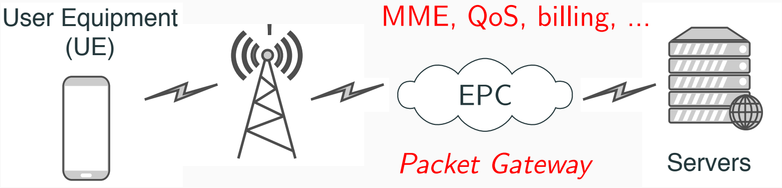
Context: Monitoring the Evolved Packet Core (EPC) in 4G

The Evolved Packet Core



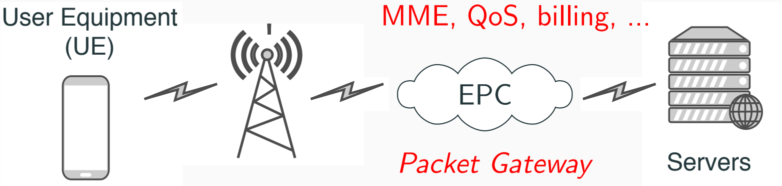
Context: Monitoring the Evolved Packet Core (EPC) in 4G

The Evolved Packet Core



Context: Monitoring the Evolved Packet Core (EPC) in 4G

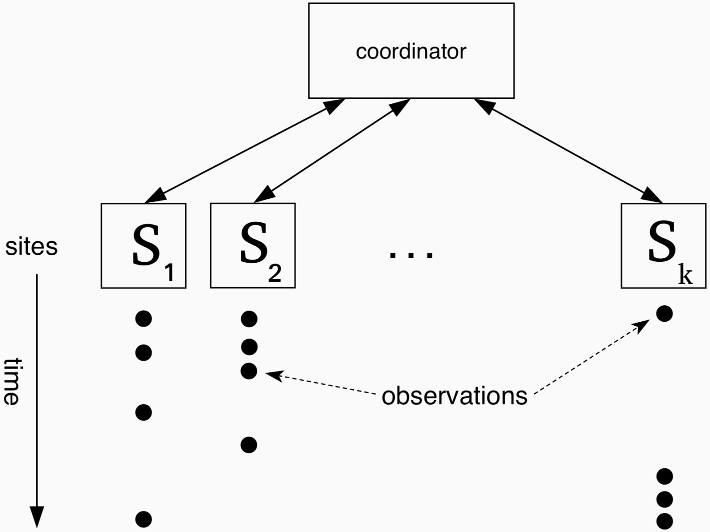
The Evolved Packet Core



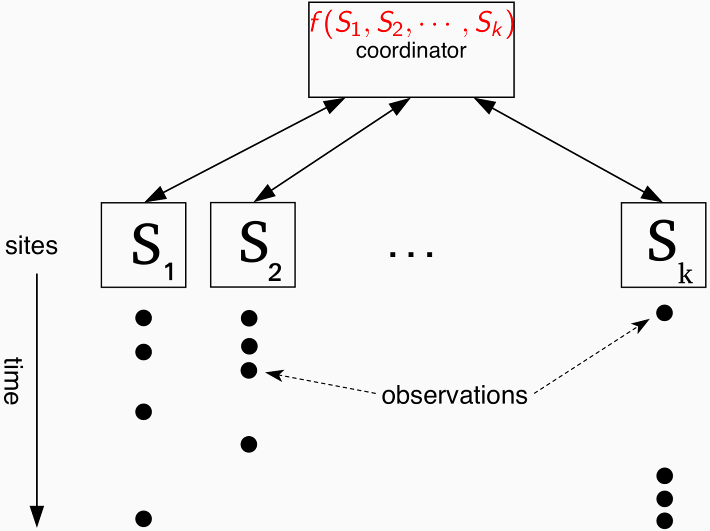
- Large-Scale, Distributed, Performance-critical system.
- Strong need to continuously monitor the EPC: e.g. detection of under- or over-used subcomponents.

Continuous Distributed Monitoring

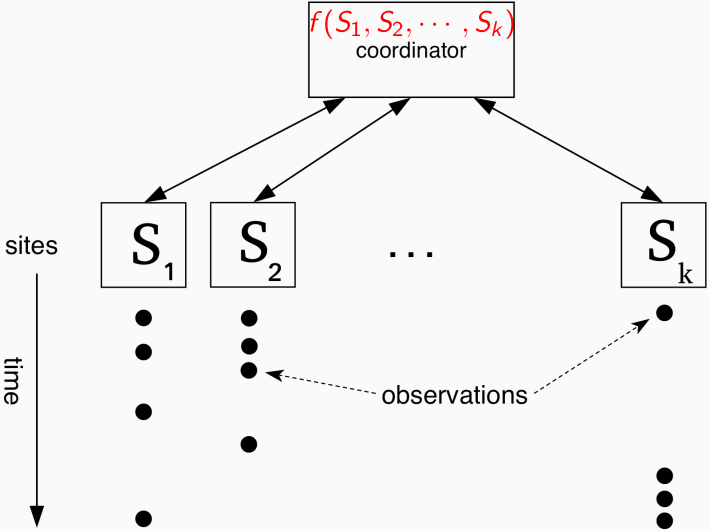
Continuous Distributed Monitoring (CDM) Model



Continuous Distributed Monitoring (CDM) Model



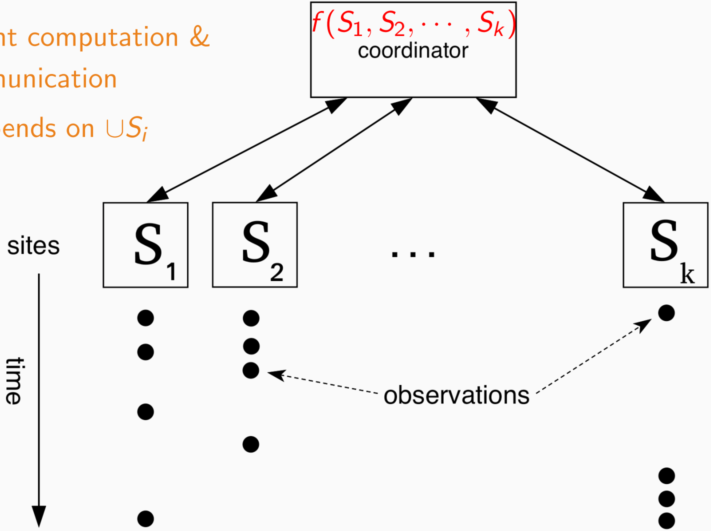
Continuous Distributed Monitoring (CDM) Model



There exist variants (unidirectional, relay nodes, etc).

Continuous Distributed Monitoring (CDM) Model

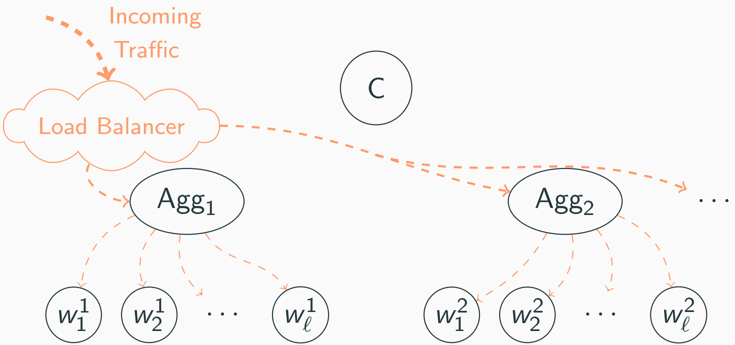
- Instant computation & communication
- f depends on $\cup S_i$



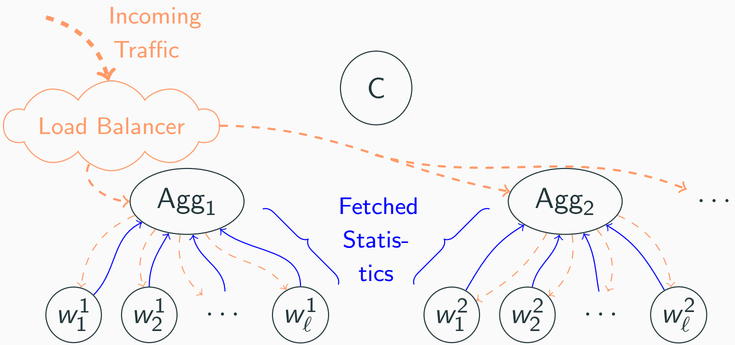
There exist variants (unidirectional, relay nodes, etc).

System Architecture

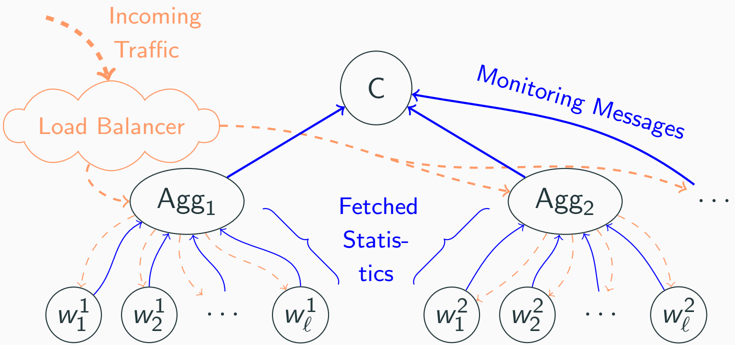
System Architecture Overview



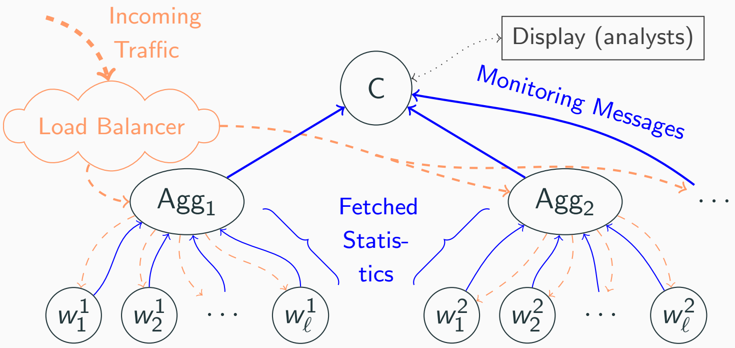
System Architecture Overview



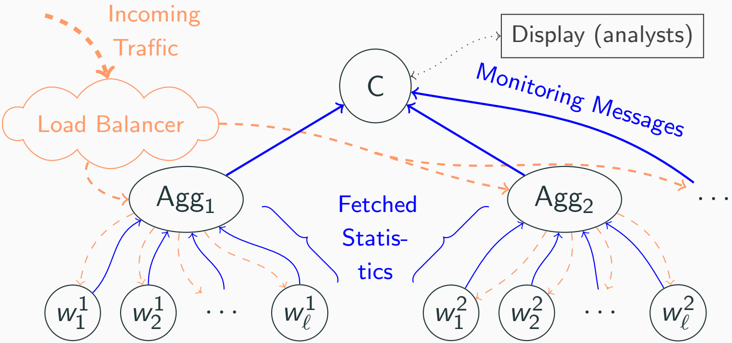
System Architecture Overview



System Architecture Overview



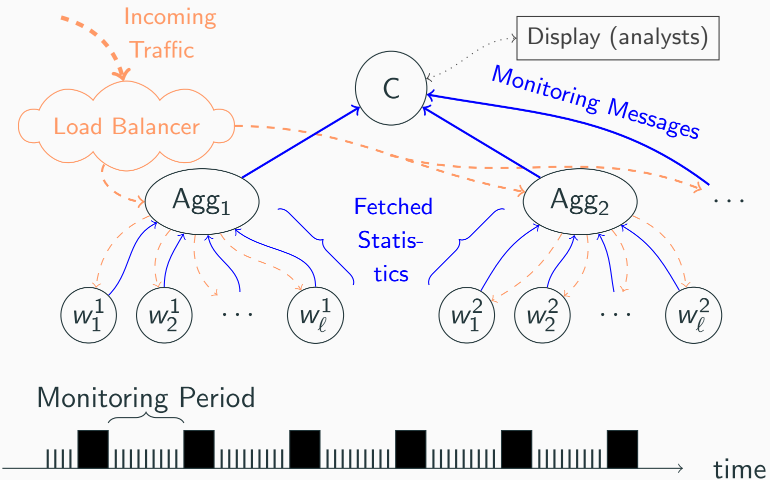
System Architecture Overview



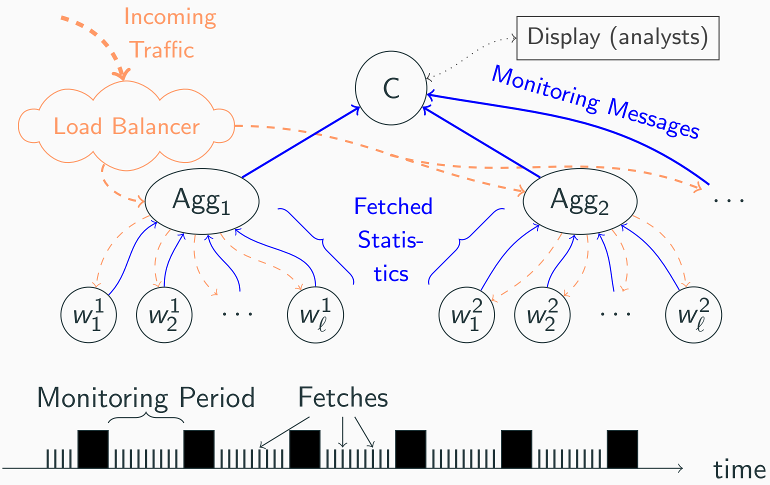
Differences with CDM models

- Sites identity matters, performance statistics $\neq$ "events", etc
- Need to account for comp. and communication delays!

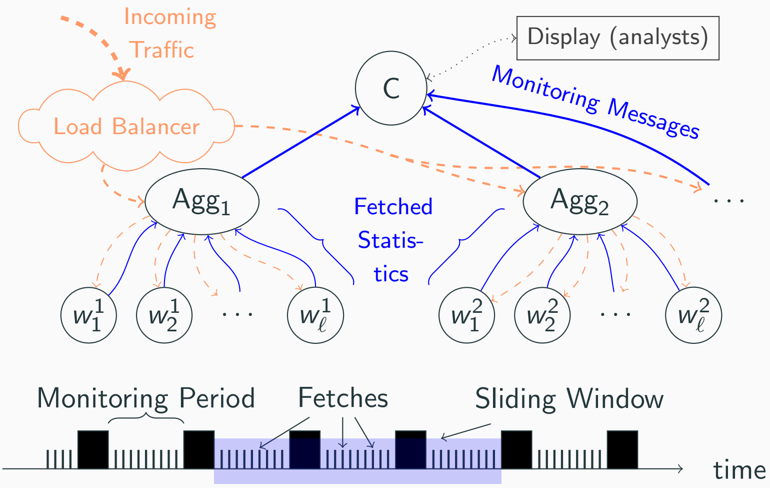
System Architecture Overview



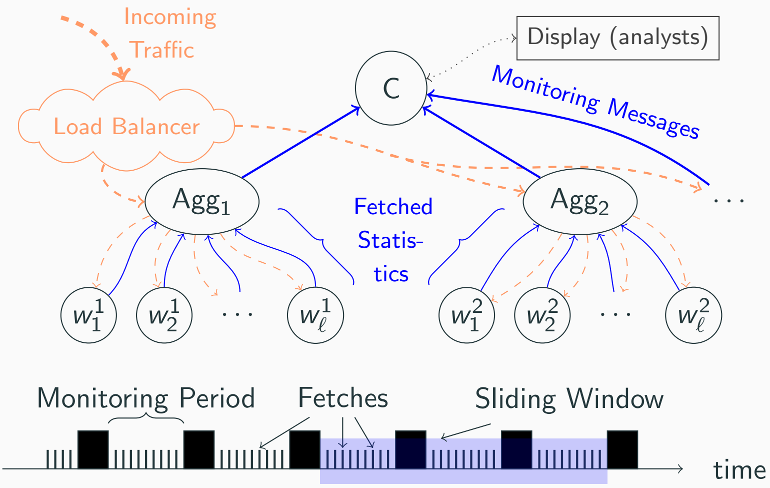
System Architecture Overview



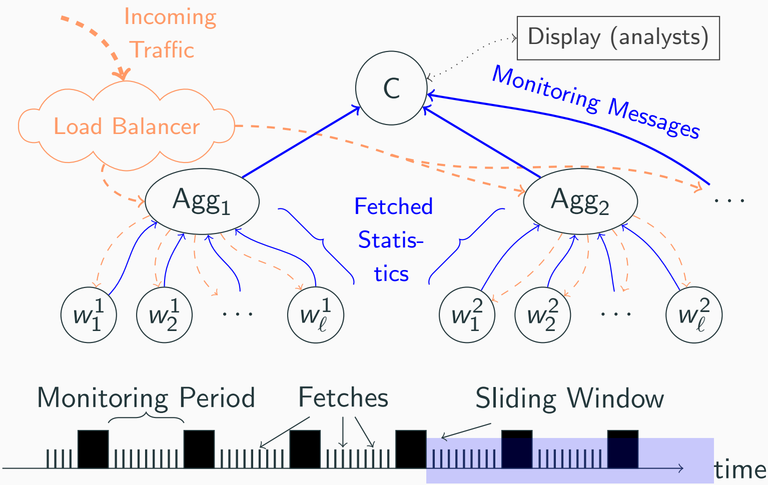
System Architecture Overview



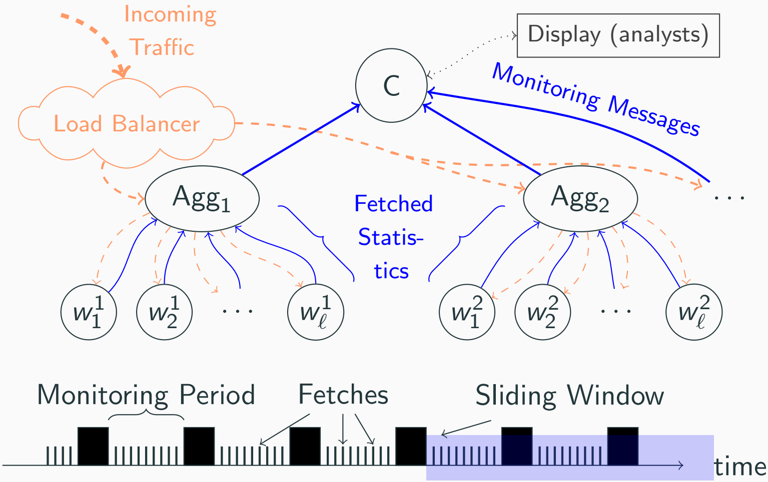
System Architecture Overview



System Architecture Overview



System Architecture Overview

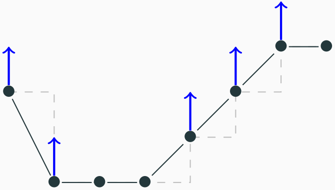


→ At the Agg: monitoring decisions then 1 monitoring message.

Monitoring Algorithms

Basic Mode: Exact Monitoring

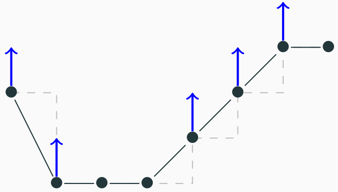
- Send an update if *last value sent* is different to measured value
- Keep an exact sliding window of the last n values



Selected CDM Algorithms for Counting problems

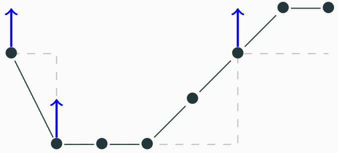
Basic Mode: Exact Monitoring

- Send an update if *last value sent* is different to measured value
- Keep an exact sliding window of the last n values



Approximation Mode: Relative Error of ε

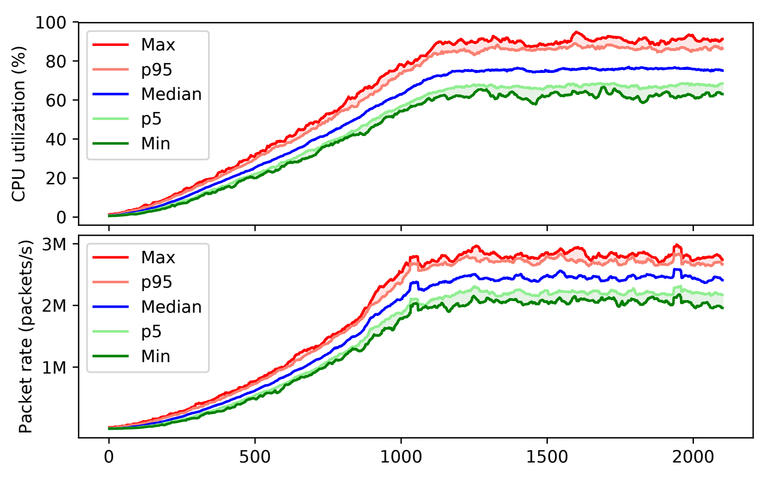
- Uses Exponential Histograms for approximate counting
- Send the *approximate count* when it is beyond some error bound from the last value sent
- Requires in all $\mathcal{O}(\log(n\varepsilon)/\varepsilon)$ words



Results

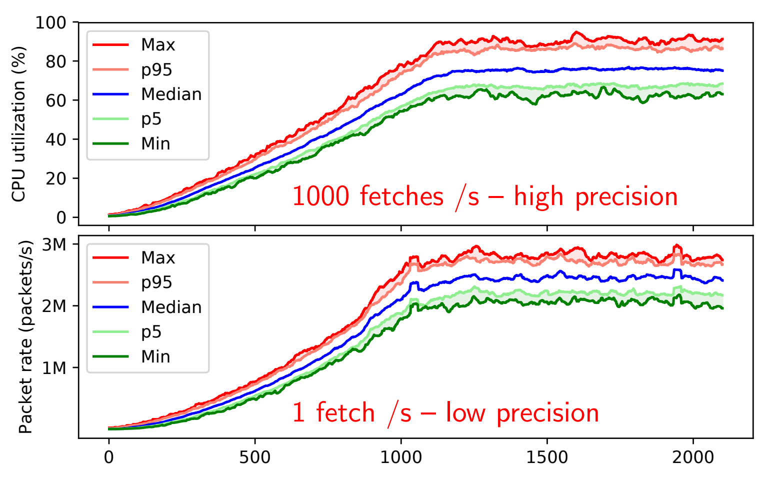
Experimental setup

- EPG setup: 2 aggregators, 72 workers per aggregator
- 2 phases: increasing load (20min) then stable load (15min)



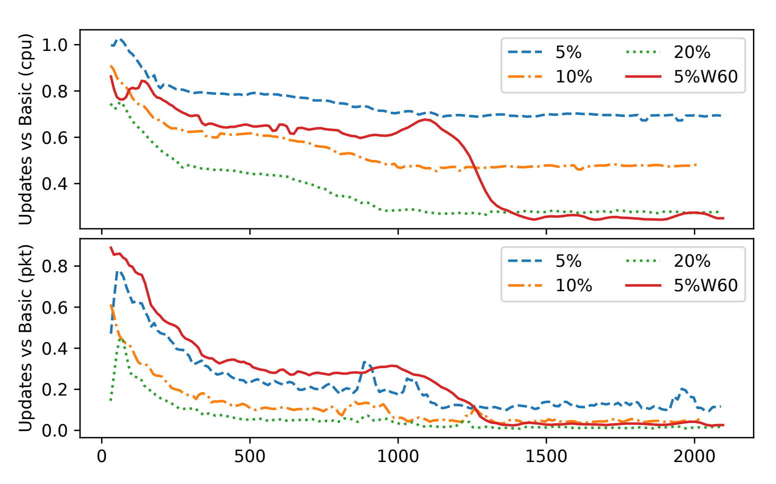
Experimental setup

- EPG setup: 2 aggregators, 72 workers per aggregator
- 2 phases: increasing load (20min) then stable load (15min)



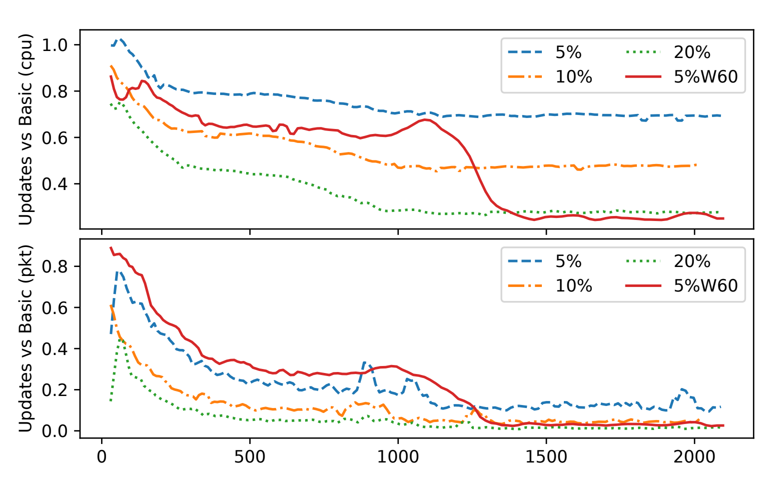
No. of Monitoring Updates per Round

- 5-10% of data sent for packet proc. rate; 30-70% for CPU.



No. of Monitoring Updates per Round

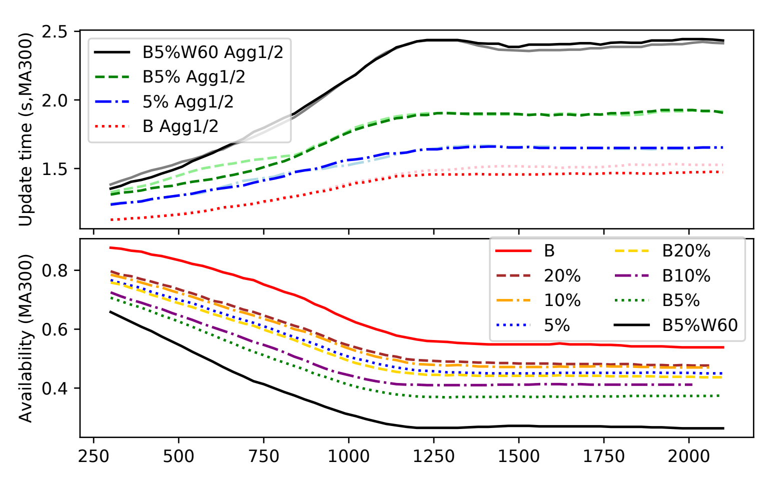
- 5-10% of data sent for packet proc. rate; 30-70% for CPU.



- Max relative error $< \frac{5\varepsilon}{9}$ and average $< \frac{\varepsilon}{5}$.

Monitoring Availability

- 8 runs (ca 4h of data) with monitoring round = 1s



Conclusion

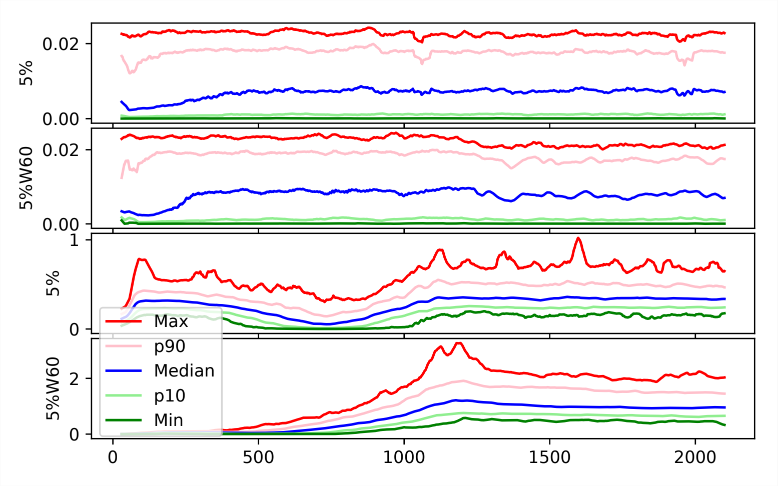
Conclusions

- Adjusted state-of-the-art CDM implementations in the EPC
- Keys to popularize CDM within a production level system
- From experiments, only 6% of data sent for 1.6% avg error
- Useful for the upcoming transition to 5G architecture

Thank you!

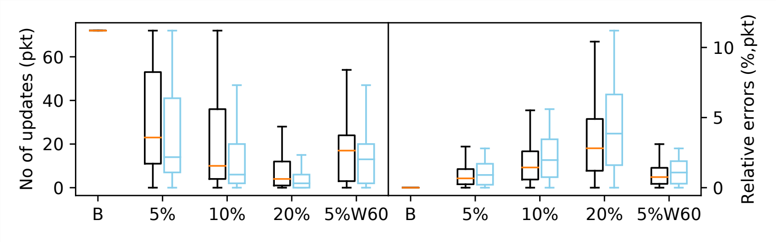
Error Analysis

- Max relative error is always close to $\frac{5\varepsilon}{9}$
- Larger window influences absolute error on CPU



Comparison with Simple Approximation

- Simple Approximation: keep an exact window and send updates when last count is beyond some predefined relative bound



- ε-Approximate algorithm presents similar tradeoffs as the simple approximation with bound $\frac{5\varepsilon}{9}$

CDM approaches

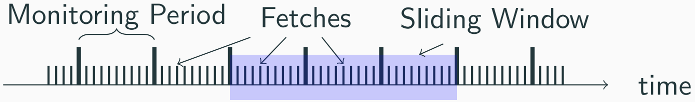
Simple approaches

- Flooding, do not scale!
- Polling, but hard to choose right polling interval!
- Sampling, do not capture scarce under/over-used components!

Solutions

- Communication-optimal algorithms
- Geometric Monitoring → efficient network-wide aggregate.
- Tailored algorithms for particular tasks → e.g. computing the frequency of items or most popular ones.
- Heuristics → e.g. adaptive filters.
- Compromises: Magpie, Dapper, Ganglia...

Proposed Monitoring Solutions



Monitoring Logic for each monitored value

- Implemented as part of the aggregator nodes
- once all fetched have been collected, a monitoring decision is taken upon propagating the update
- Aggregation of all monitoring updates: sending of (up to) a single monitoring message per aggregator

Selected CDM Algorithms

Basic Mode

- Send an update if last value sent is different
- Keep an exact sliding window of length n

ε-Approximation Mode

- Maintains an $\frac{\varepsilon}{9}$-approximate Exponential Histogram for counting approximate sum $\hat{c}$ of items over a sliding window of the last n events
- Whenever $\hat{c} > (1 + \frac{4\varepsilon}{9})c$ or $\hat{c} < (1 - \frac{4\varepsilon}{9})c$, send an update, where c is the last value sent
- Requires in all $\mathcal{O}(\log(n\varepsilon)/\varepsilon)$ words of memory

Measuring Metrics of Interests: 2 modes

With high granularity: CPU usage

1. P fetches of CPU-usage for past 1ms each within one monitoring period
2. Frequency chart (histogram of F bins) for the P fetches
3. *Sliding Windows are updated*: each bin is monitored
4. For each changed (basic) or outside of bounds (approx) value, a monitoring update is sent
5. Upon receiving an update: C updates its frequency counts for the resp. observer and CPU-bin and then may display the average CPU over the window as $\sum_{1 \leq i \leq F} i f_i / \sum_{1 \leq i \leq F} f_i$

With low granularity: Packet Processing Rate

- Only the no. of processed packets per mon. period is tracked