Brief Announcement:
Tracking Distributed Aggregates over Time-based Sliding Windows

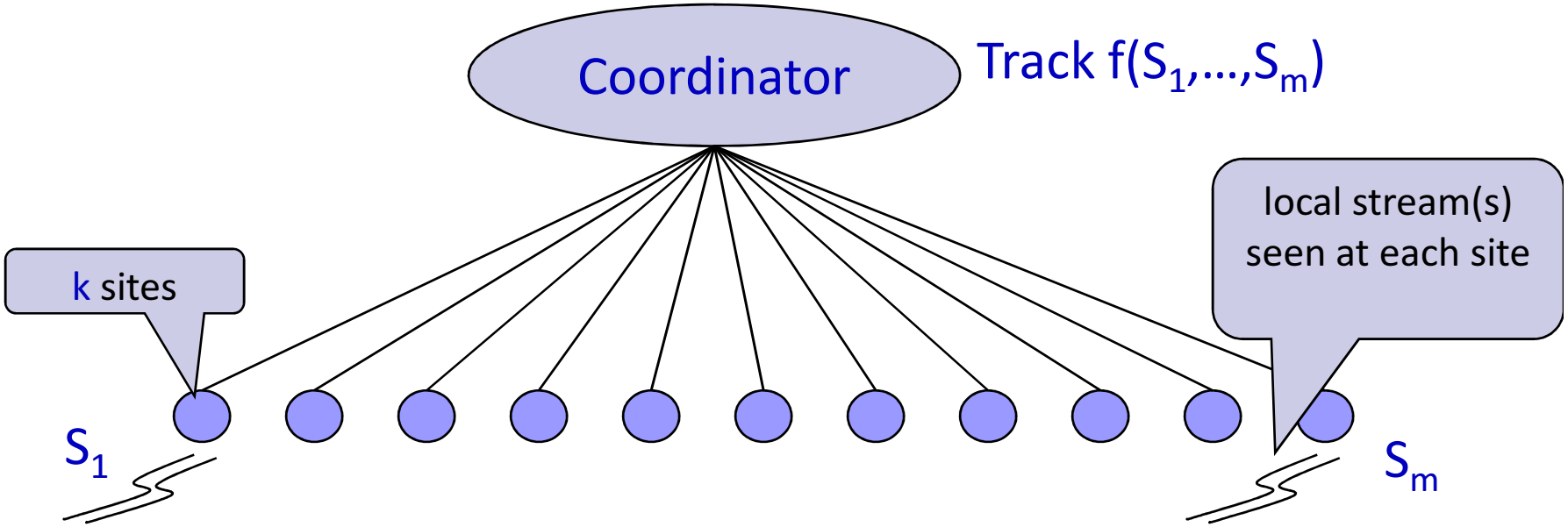
Graham Cormode

AT&T Labs

Ke Yi

HKUST

Continuous Distributed Model

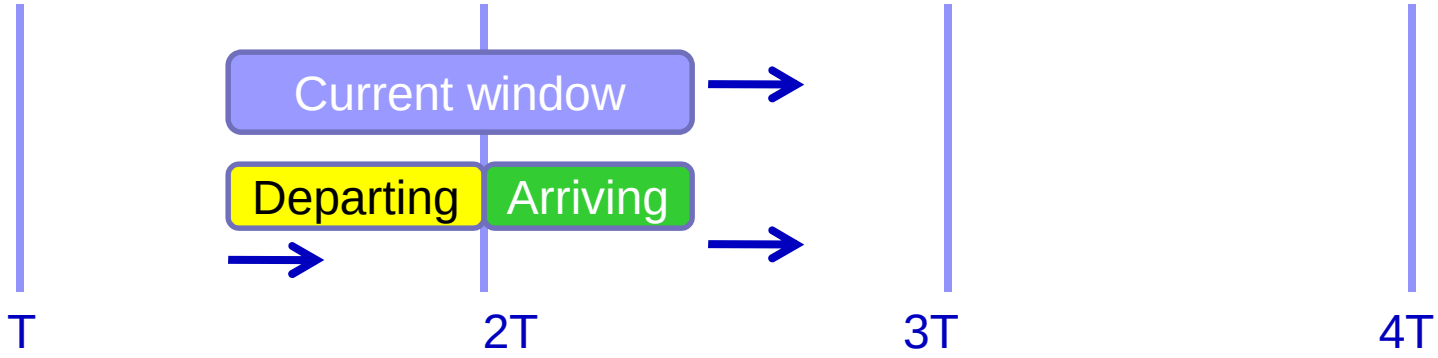


- Other structures possible (e.g., hierarchical)
- Site-site communication only changes things by factor 2
- **Goal**: Coordinator *continuously tracks* (global) function of streams
  - Achieve communication and space poly(k,1/ε,log n)

Problems in Distributed Monitoring

- Much interest in these problems in TCS and Database areas
- Track holistic functions of the (global) data distribution
  - Quantiles and heavy hitters [C, Garofalakis, Muthukrishnan, Rastogi 05]
  - Empirical Entropy [Arackaparambil Brody Chakrabarti 09]
  - Frequency Moments [C, Muthukrishnan, Yi 08]
  - Geometric approach [Sharman, Schuster, Keren 06]
- Track functions only over sliding window of recent events
  - Samples [C, Muthukrishnan, Yi, Zhang 10]
  - Counts and frequencies [Chan Lam Lee Ting 10]
- This work: new framework for monitoring over sliding windows

Forward/backward framework



- Key insight:
  - Complexity of sliding window comes from non-monotonicity
  - Break any window into forward (arrivals) and backward (expiries)
  - Solve each separately, improving overall
- Optimal results for several problems follow easily
  - Counting: $O(k/\varepsilon \log (\varepsilon n/k))$ communication, $O(1/\varepsilon \log \varepsilon n)$ space
  - Heavy hitters: $O(k/\varepsilon \log (\varepsilon n/k))$ communication, $O(1/\varepsilon \log \varepsilon n)$ space
  - Quantiles: $O(k/\varepsilon \log^2 1/\varepsilon \log (\varepsilon n/k))$ comm, $O(1/\varepsilon \log^2 1/\varepsilon \log \varepsilon n)$ space