Distributed Online Tracking

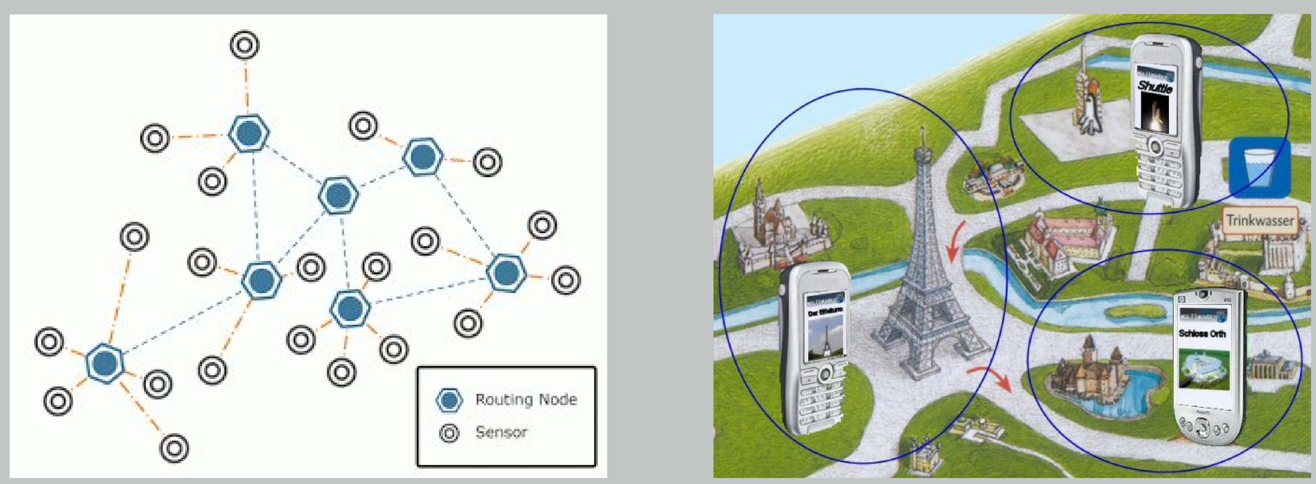
Mingwang Tang[1], Feifei Li[1], Yufei Tao[2]

[1]University of Utah
{tang, lifeifei}@cs.utah.edu

[2]Chinese University of Hong Kong
{taoyf}@cse.cuhk.edu.hk

Motivation

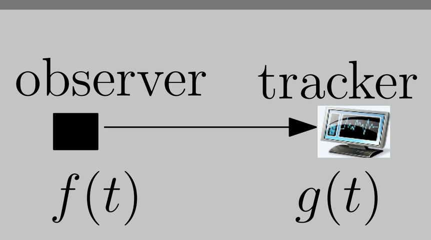
- Tracking a user function over distributed data in online fashion is a fundamental challenge.

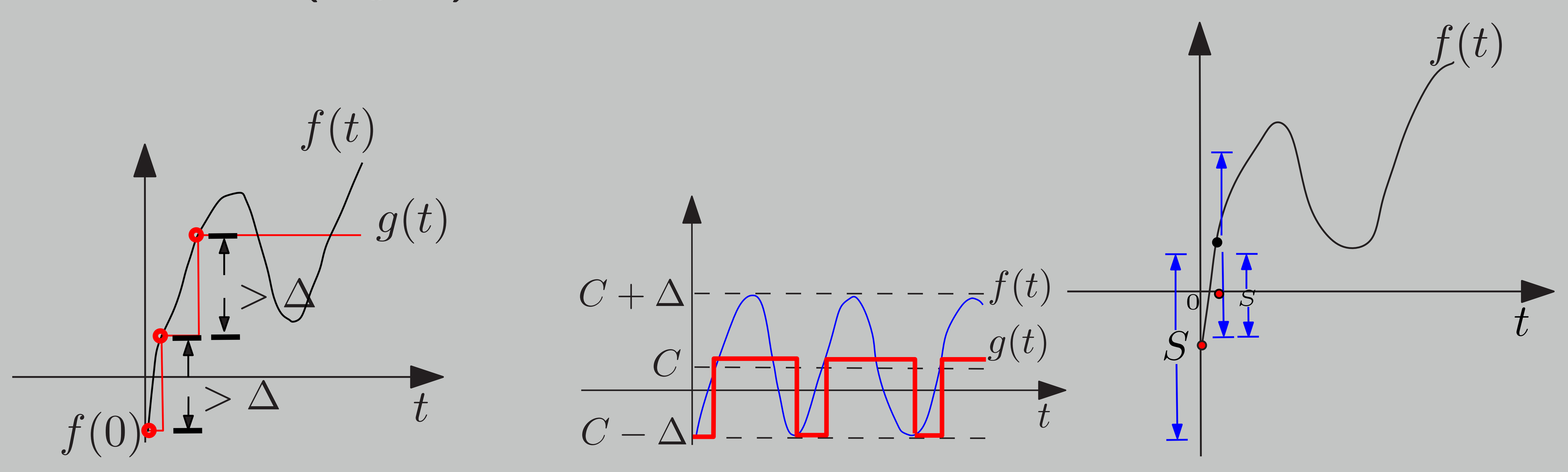


(a) sensor network. (b) location based service.

Background: two-party online tracking

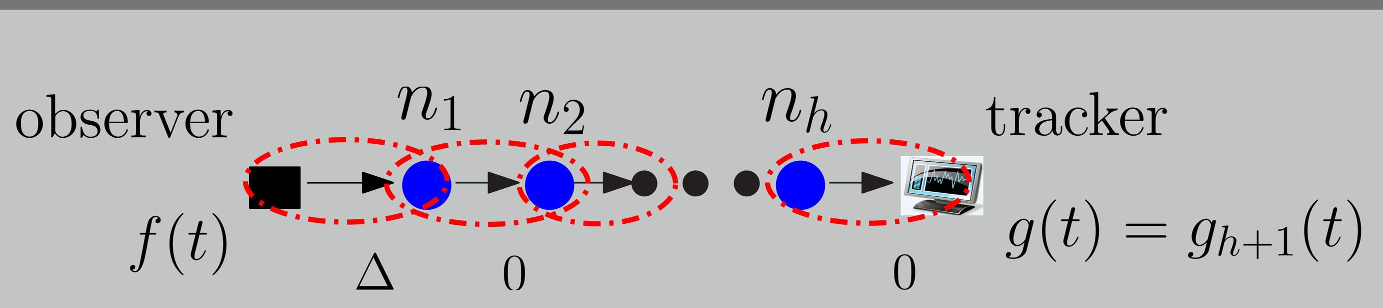
- $f: \mathbb{Z} \to \mathbb{Z}$; $g(t) \in [f(t) - \Delta, f(t) + \Delta]$.
- Naive method: unbounded competitive ratio; OptTrack: $O(\log \Delta)$ competitive ratio.





(a,b) Naive method. (c) OptTrack method.

Chain online tracking



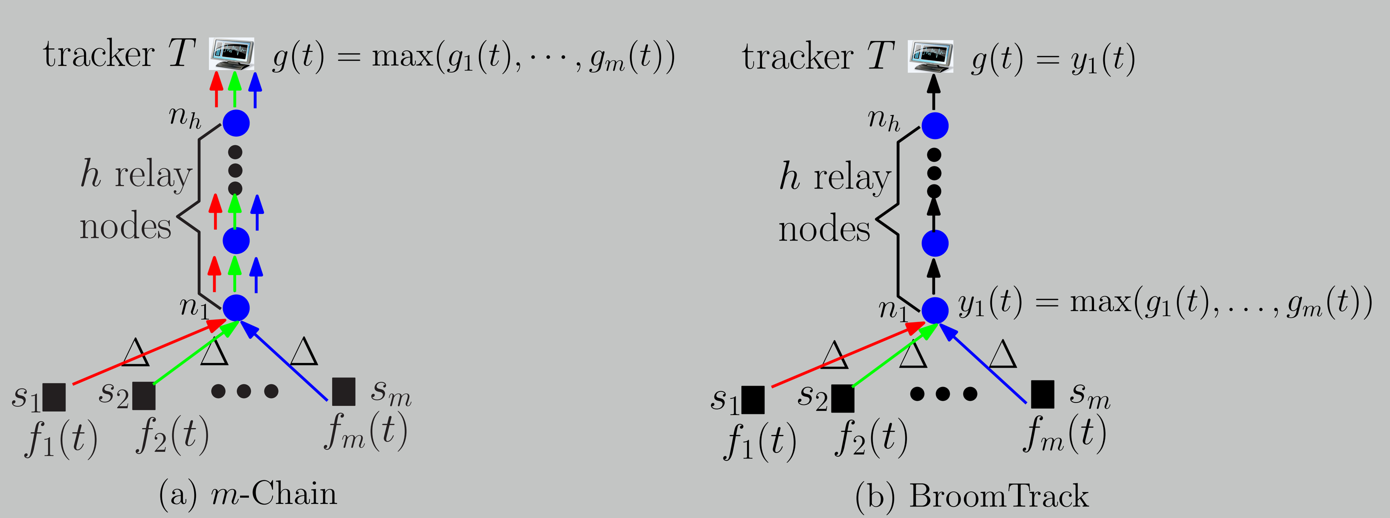
- ChainTrackA: distribute Δ averagely.
- ChainTrackO: distribute Δ randomly among $h+1$ centralized instances.
- ChainTrackO: assign the whole tracking error Δ to the first tracking instance. It achieves $O(\log \Delta)$ competitive ratio.

Distributed setting

- Each observer s_i observes an arbitrary function f_i over time.
- T wants to keep tracking of $f(t) = f(f_1(t), f_2(t), \ldots, f_m(t))$ using $g(t)$ within an error Δ for any time instance t.

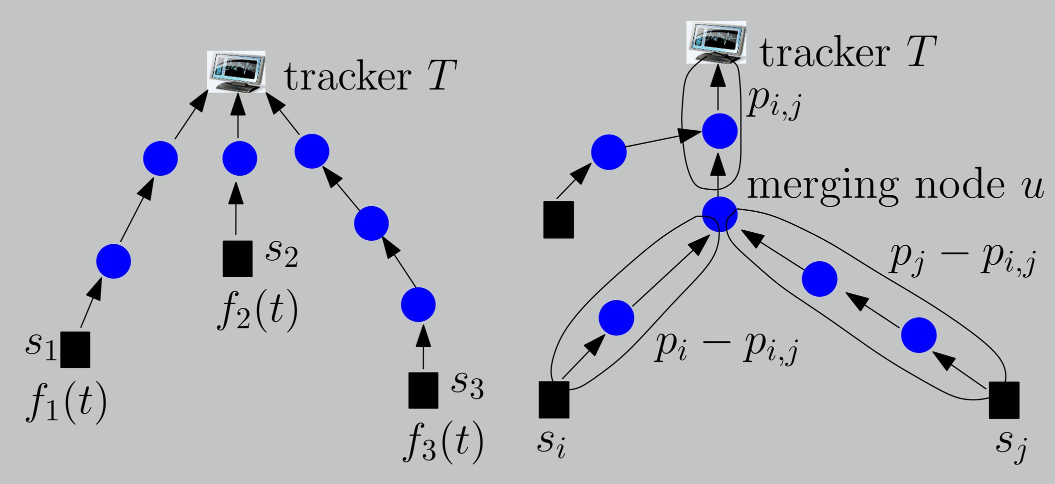
Broom online tracking: max

$f(t) = \max(f_1(t), \cdots, f_m(t))$ $\quad g(t) \in [f(t) - \Delta, f(t) + \Delta]$



- Theorem 1: For any algorithm A in A_{broom}, there exists an input instance I and another algorithm $A' \in A_{broom}$, such that cost(A, I) is at least h times worse than cost(A', I), i.e., for any $A \in A_{broom}$, ratio$(A) = \Omega(h)$.
- Theorem 2: With respect to online algorithms in A_{broom}, ratio(BroomTrack) $< h \log \Delta$.
- Corollary 1: ratio(m-Chain) $= O(h \log \Delta)$.

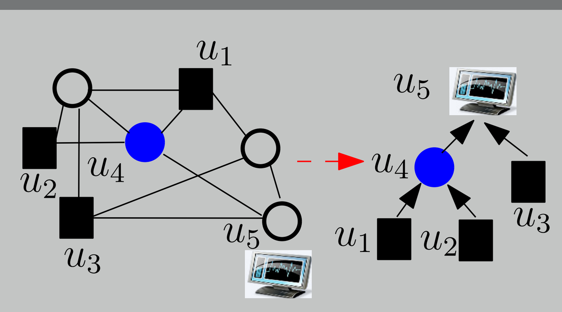
General tree online tracking: max



(a) simple tree. (b) general tree

- Corollary 2: There is no instance optimal algorithm for A_{tree}.
- Corollary 3: ratio(TreeTrack) $= O(h_{max} \log \Delta)$ with respect to A_{tree}.

Other topologies and functions



graph topology

- Other functions: min, sum, average and ϕ-quantile

Experiment

- Temperature (TEMP) from MesoWest Project.
- Wind Direction (WD) from SAMOS project.

$N = 500$	# time instances
$h = 2$	# relay nodes
$f = max$	aggregate function
$\Delta = 0.6\tau$	error threshold, $\tau = avg(std(f_1), \ldots, std(f_m))$
$m = 15$	# observers

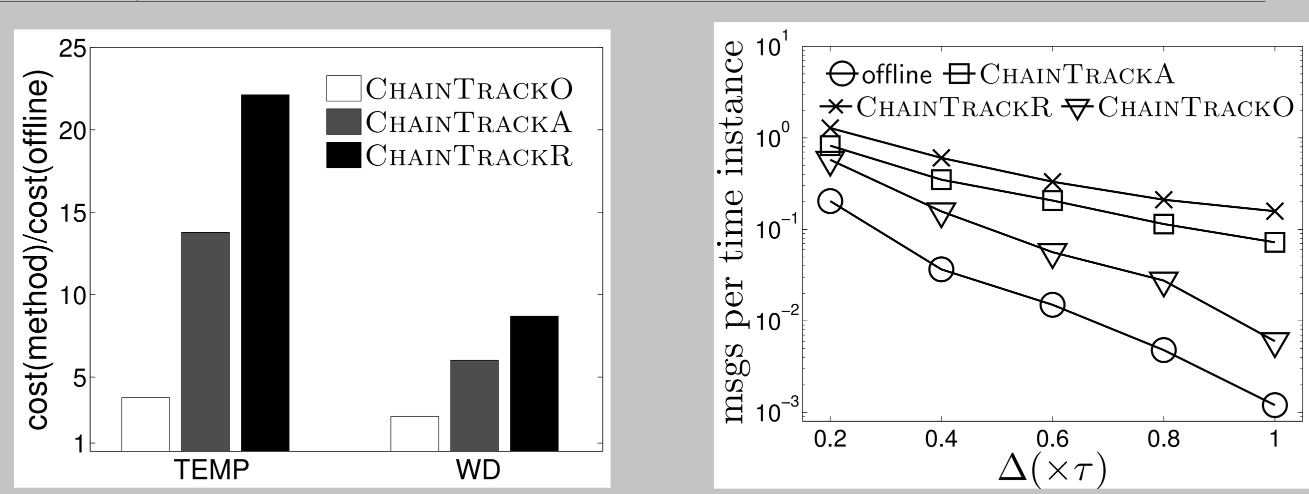


Figure 1: Chain online tracking

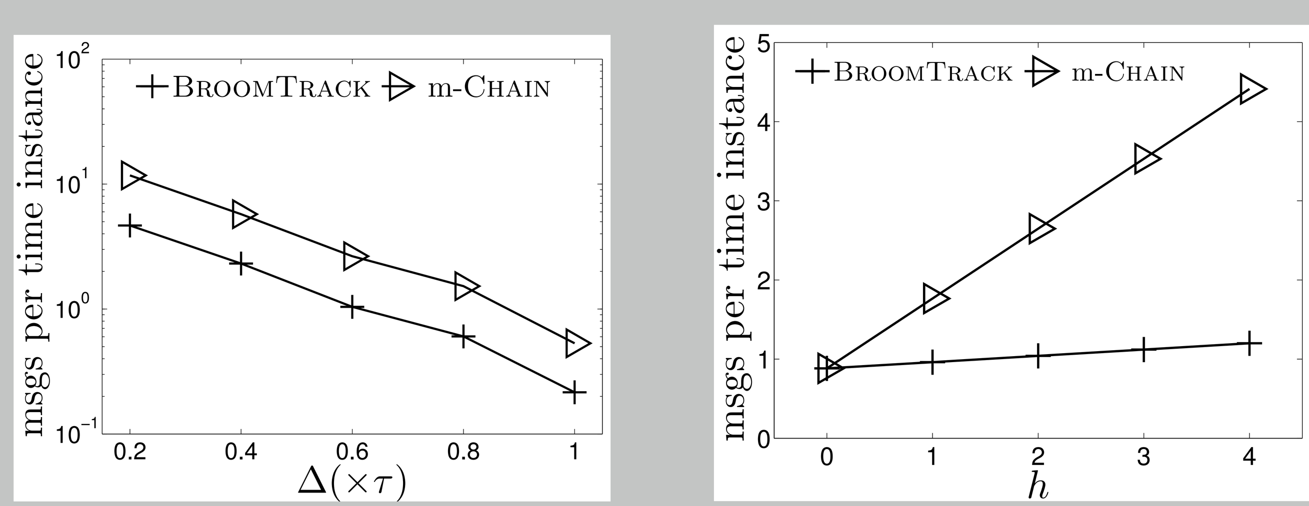


Figure 2: Broom online tracking

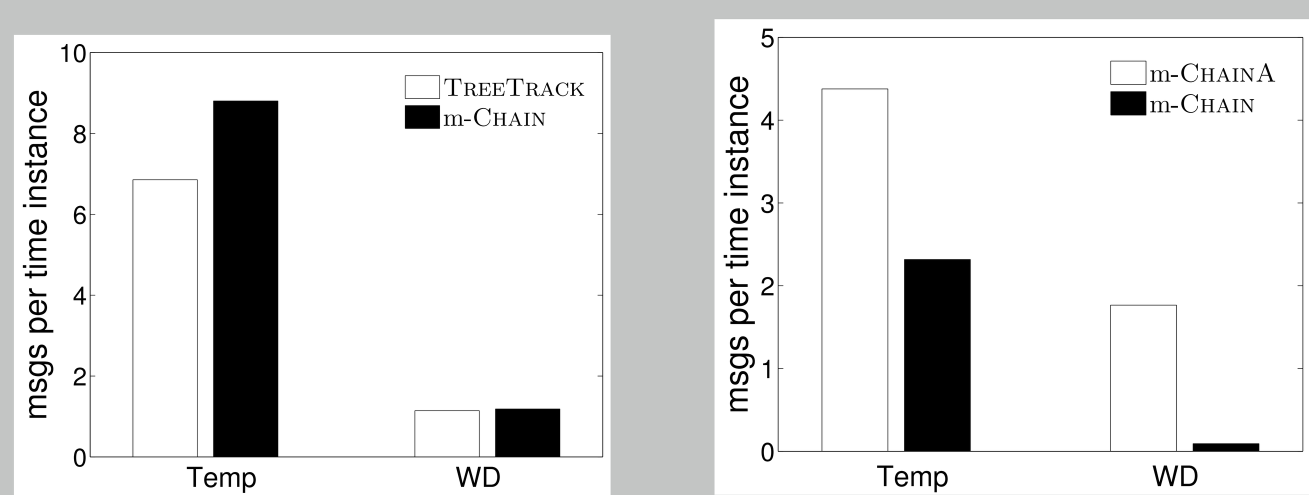


Figure 3: (a) sum on general tree. (b) median on general tree

Conclusion

- Extend the centralized, two party model to the chain model.
- Investigate both the broom model and the tree model, as well as other different tracking functions.