Online Minimum Matching in Real-Time Spatial Data: Experiments and Analysis

Yongxin Tong [1], Jieying She [2], Bolin Ding [3], Lei Chen [2], Tianyu Wo [1], Ke Xu [1]

[1] SKLSDE Lab, NSTR, and IRI, Beihang University, China

[2] The Hong Kong University of Science and Technology, Hong Kong, China

[3] Microsoft Research, Redmond, WA, USA

[1] {yxtong, woty, kexu}@buaa.edu.cn, [2] {jshe, leichen}@cse.ust.hk, [3] bolind@microsoft.com

Introduction

- Online Minimum Bipartite Matching in Spatial Data (OMBM)
- Most applications of OMBM need to be addressed in real-time
  - Task Assignment in Spatial Crowdsourcing
  - Taxi Dispatching
  - Food Delivery
- Motivations and Contributions

	Motivations	Contributions
1	Is Greedy really the worst?	Greedy has good performance.
2	Is the worst-case analysis appropriate for the OMBM problem in practice?	Worst-case vs. Average-case analysis.
3	Are implementations and experimental evaluations uniform?	Uniform implementations and evaluations are provided.

The OMBM Problem

- Given
  - A set of (**static**) service providers W
    - Each $t \in T$: location l_t.
  - A set of (**dynamic**) users T
    - Each $w \in W$: location l_w and arriving time a_w.
  - Cost Function $dis(t,w)$: any metric distance function
- Find an online matching M to minimize the total cost $MinSum(M)=\sum_{t\in T, w\in W} dis(t,w)$ s.t.
  - Cardinality Constraint: $|M|=\min\{|T|,|W|\}$
  - Real-Time Constraint: Once a user t appears, a service provider must be immediately assigned to t before the next user appears.
  - Invariable Constraint: Once a user t is assigned to a service provider w, the match (t, w) cannot be changed
- Online Algorithm Evaluation: Competitive Ratio (CR)
  - Adversarial Model: Worst-Case Analysis
    - $CR_A = max_{\forall G(T,W,U) and \forall v \in V} \frac{MinSum(M)}{MinSum(OPT)}$
  - Radom Order Model: Average-Case Analysis
    - $CR_{RO} = max_{\forall G(T,W,U)} \frac{\mathbb{E}[MinSum(M)]}{MinSum(OPT)}$

Arrival Time	9:30	9:32	9:35
1st Order	t_1	t_2	t_3
2nd Order	t_2	t_3	t_1

The performance for an online algorithm depends on different arrival orders.

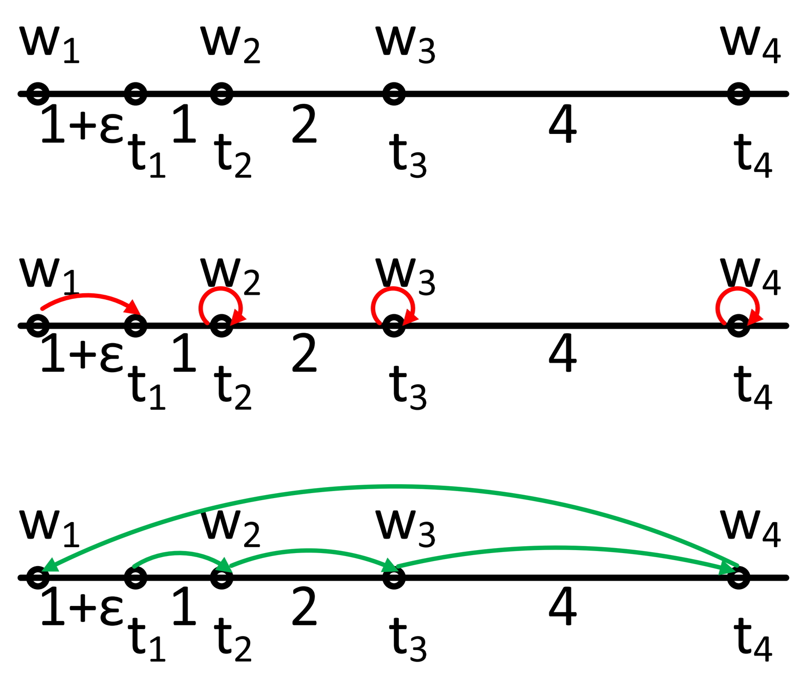
Representative Algorithms for OMBM Problem

Algorithms	Time Complexity per Each Arrival Vertex	Randomization	Data Structure	Competitive Ratio
Greedy [SODA 1991]	O(k)	Deterministic	No	$O(2^k)$
Permutation [SODA 1991]	$O(k^3)$	Deterministic	No	O(2k-1)
HST-Greedy [SODA 2006]	O(k)	Randomized	HST	$O(\log^3 k)$
HST-Reassignment [ESA 2007]	$O(k^2)$	Randomized	HST	$O(\log^2 k)$

- HST: Hierarchically Separated Tree [STOC 2003]

Is the greedy algorithm really the worst in real practice?

Greedy Algorithm Revisited

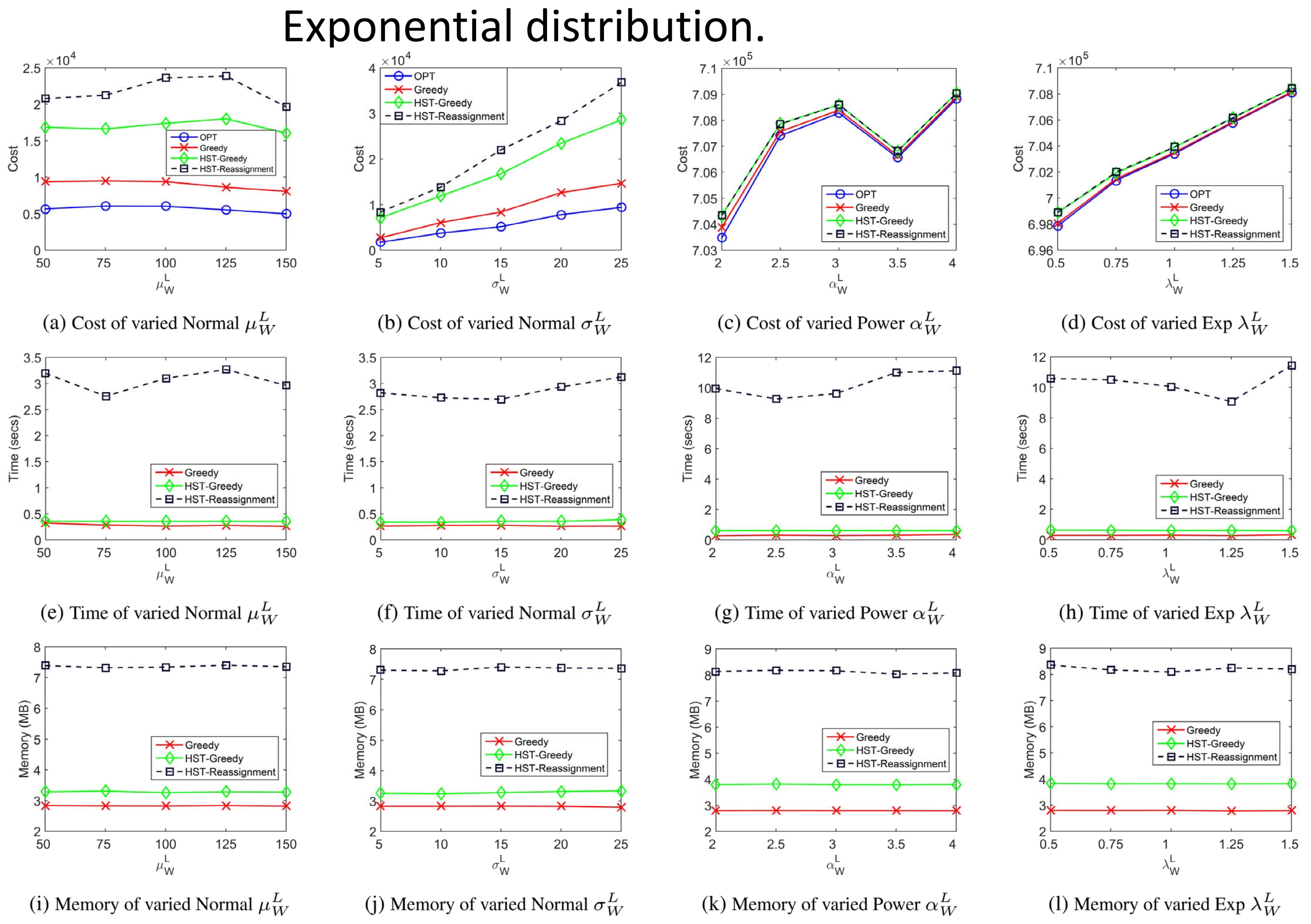


Competitive Ratio (Adversarial Model in this case): 2^k

Competitive Ratio (Random Order Model in this case): 3.195

Experimental Evaluation

- Real Datasets (Shenzhou Taxis)
  - 15082 Shenzhou Taxis at Beijing in May 2015
  - Average 115364 calling-taxi requests per day
- Synthetic Datasets (5000 ×5000 grids)
  - The locations randomly follows Normal distribution, Uniform distribution, Power-law distribution and Exponential distribution.



(a) Cost of varied Normal μ_W^L (b) Cost of varied Normal σ_W^L (c) Cost of varied Power α_W^L (d) Cost of varied Exp λ_W^L

(e) Time of varied Normal μ_W^L (f) Time of varied Normal σ_W^L (g) Time of varied Power α_W^L (h) Time of varied Exp λ_W^L

(i) Memory of varied Normal μ_W^L (j) Memory of varied Normal σ_W^L (k) Memory of varied Power α_W^L (l) Memory of varied Exp λ_W^L

Results that the locations of service providers in W follow Normal, Power-law, and Exponential distributions while the locations of users follow Normal distribution.