# Research Statement

Zheng Baihua
School of Information Systems, Singapore Management University
Tel: (65) 6828-0915; Email: bhzheng@smu.edu.sg
Updated on February 01, 2012

## Background

With the ever growing popularity of smart mobile devices and rapid advent of wireless technology, the vision of ubiquitous computing has come closer to reality. Ubicomp devices are mostly driven by new generation pervasive or context-aware information services. For example, location-based services provide information relevant to users' geographical location. However, wider acceptance of such services has been hindered by issues from scalability, privacy to mobile devices' limited power resources, memory spaces and computation capability.

My research focuses on developing data management techniques in support of efficient processing of complex queries. My major contributions are in the area of location-dependent queries, applicable to a wide range of mobile environments including wireless broadcast systems, mobile ad-hoc network, and wireless sensor network. My research outputs can be further divided into three smaller areas, namely data indexing, client side caching, and server side scheduling.

## Research Areas

Data Indexing: I have been working on data indexing techniques to support location-dependent queries since my PhD study. Initially, I focused on the dissemination of location-dependent data in wireless broadcast systems. The fact that the data broadcast on air are only periodically available renders most existing indexing structures inefficient if not useless, as they were designed for traditional database systems and their searching capabilities depend on random access and backtracking. Taking linearity requirement into account, I designed several new indexes to support various location-dependent queries, such as point queries, window queries, KNN queries, and continuous-KNN queries, in wireless broadcast systems. I further extended my initial study in two directions. First, instead of purely focusing on the theoretical design of index structures, I started to explore the properties of wireless channels to ensure performance in real-world systems. For example, wireless channels by nature are error-prone and occasional data packet loss is a given. I designed a distributed index with multiple embedded search paths to allow the users to switch to another path in case the current path is lost. Second, I formulated and studied new types of queries motivated by real applications. For example, nearest surrounder queries consider not only the proximity but also the angular relationship between objects and query points.

In addition to wireless broadcast systems, I have also developed high performance index structures for location-dependent query processing in mobile systems of different properties and requirements, such as road networks and sensor networks. Recently, I have been

working on the index structures for complex query processing in obstructed spaces. This is motivated by applications where neither Euclidean distance nor network distance can represent the proximity between objects. For example, battlefields usually have no fixed road network structure and soldiers would enjoy certain degree of free movement until they hit obstacles. To this end, I have designed efficient index structures for common location-dependent queries (e.g., continuous-KNN queries and reversed KNN search).

Client side caching: Clients cache data in their local memory while moving in a mobile environment. Consequently, it is beneficial to understand how we can fully utilize local cached data. My work has addressed two fundamental issues related to mobile client cache: i) the client cache may become invalid due to the change of its location; and ii) new queries may be issued as client moves around. To ensure client cache validity, I developed a method to cache not only data object but also corresponding valid region, i.e., a geographical region within which the cached object is guaranteed to be valid for the same query. I also investigated different presentations of valid scopes.

Server side scheduling: Given the fact that mobile users have different requests and some data objects are more popular than the others, scheduling algorithms try to rearrange data objects to minimize the access time, reduce power consumption, and optimize bandwidth utilization. I studied the scheduling algorithm in a multi-channel environment where data objects of various sizes and access frequencies are broadcast in parallel on multiple channels of different bandwidths, with the client listening to one channel at a time. I also raised a privacy concern regarding the scheduling algorithms as most of the existing scheduling algorithms require the clients to reveal their access patterns to the server. A high-performance privacy preserving scheduling algorithm was proposed to make available the overall access pattern of data objects without disclosing individual access pattern.

My ongoing research centers around above areas, and meanwhile I am investigating data management related issues in social network. For example, Twitter is a highly popular microblogging site where hundreds of millions of users follow one another for information sharing and interaction. Given the vast amount of realtime Twitter data, constrained data crawling bandwidth, and the need for crawled data to be meaningful for problem-specific data analytics research, we need to define appropriate target scopes for data crawling.

## Selected Publications and Outputs

Zhisheng Li, Ken C.K. Lee, **Baihua Zheng**, Wang-chien LEE, Dik Lun LEE, and Xufa WANG: IR-tree: An Efficient Index for Geographic Document Search", *IEEE Transactions on Knowledge and Data Engineering (TKDE),* 23(4)*:* 585-599, April 2011

Xi Guo, **Baihua Zheng**, Yoshiharu ISHIKAWa, and Yunjun Gao: Direction-Based Surrounder Queries for Mobile Recommendations, *VLDB Journal*, 29(5): 743-766, October 2011

Yunjun Gao, **Baihua Zheng**, Gang Chen, Chun Chen, and Qing Li: Continuous Nearest Neighbor Search in the Presence of Obstacles, *ACM Transactions on Database Systems (TODS),* 36(2), May 2011

Ken C. K. Lee, Brandon Unger, **Baihua Zheng**, and Wang-chien Lee: Location-Dependent Spatial Query

Containment, *Data and Knowlegde Engineering (DKE)*, 70(10): 842-865, October 2011

Yunjun Gao, **Baihua Zheng**, GenCai Chen, Qing Li, and Xiaofa Guo: Continuous Visible Nearest Neighbor Query Processing in Spatial Databases", *VLDB Journal*, 20(3): 371-396, June 2011

HweeHwa Pang, Xuhua Ding, and **Baihua Zheng**: Efficient Processing of Exact Top-*k* Queries over Disk-Resident Sorted Lists. *VLDB Journal*, 19(3): 437-456, June 2010.

Ken C. K. Lee, Wang-chien Lee, **Baihua Zheng**, Huajing Li and Yuan Tian: Z-SKY: An Efficient Skyline Query Processing Framework Based on Z-Order. *VLDB Journal*, 19(3): 333-362, June 2010.

Ken C. K. Lee, Wang-Chien Lee, Hong Va Leong, and **Baihua Zheng**: Navigational path privacy protection: navigational path privacy protection. *Proc. the ACM 18th Conference on Information and Knowledge Management (CIKM'09)*. Nov 2-6, 2009, Hong Kong: 691-700

Yunjun Gao and **Baihua Zheng**: Continuous Obstructed Nearest Neighbor Queries in Spatial Databases. *Proc. the ACM SIGMOD International Conference on Management of Data (Sigmod'09)*, June 29- July 02, 2009 Providence, Rhode Island, USA: 577-590

**Baihua Zheng**, Wang-chien Lee, Peng Liu, Dik Lun Lee and Xuhua Ding: Tuning On-Air Signatures for Balancing Performance and Confidentiality. *IEEE Trans. Knowledge and Data Engineering (TKDE)*, 21(12): 1783-1797, 2009.

**Baihua Zheng**, Wang-chien Lee, Ken C. K. Lee, Dik Lun Lee and Min Shao: A Distributed Spatial Index for Error-Prone Wireless Data Broadcast. *VLDB Journal*, 18(4): 959-986, 2009.

Yunjun Gao, **Baihua Zheng**, Gencai Chen, Wang-chien Lee, Ken C. K. Lee and Qing Li: On Efficient Visible Reverse k-Nearest Neighbor Query Processing in Spatial Databases. *IEEE Trans. Knowledge and Data Engineering (TKDE)*, 21(9): 1314-1327, 2009

Yunjun Gao, **Baihua Zheng**, Gencai Chen, and Qing Li: On Efficient Mutual Nearest Neighbor Query Processing in Spatial Databases. *Data & Knowledge Engineering (DKE)*, 68(8): 705-727, 2009

Yunjun Gao, **Baihua Zheng**, Gencai Chen, and Qing Li: Optimal-Location-Selection Query Processing in Spatial Databases. *IEEE Trans. Knowledge and Data Engineering (TKDE)*, 21(8):1162-1177, 2009

Ken C. K. Lee, **Baihua Zheng,** and Wang-chien Lee: Ranked Reverse Nearest Neighbor Search. *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 20(7): 894-910, 2008