# A P2P Simulation Model to Support Mobile, Scalable Nearest Neighbor Queries for Location-based Services*

Wei-Shinn Ku, Roger Zimmermann, Chi-Ngai Wan and Haojun Wang
Computer Science Department, University of Southern California, Los Angeles, CA 90089
`[wku, rzimmerm, cwan, haojunwa]@usc.edu`

## Abstract

*With the increasing capabilities of mobile devices, there has been a growing interest in location-based services. Here we present MAPLE, a scalable peer-to-peer nearest neighbor (NN) query simulation system for mobile environments. MAPLE is designed for the efficient sharing of query results cached in the local storage of mobile peers. The most innovative aspect of the MAPLE model is its ability to either fully or partially compute location-dependent nearest neighbor objects on each host. Our novel verification algorithms enable a mobile host to locally determine whether the points of interest obtained from its peers are relevant to its own NN query solution set. Our simulation system emulates all aspects of the MAPLE model including the movement of mobile hosts on a road network and the launch and execution of location-based nearest neighbor queries. The simulation illustrates how cooperative data sharing and distributed processing among mobile peers results in a considerable reduction of the load on remote spatial database servers. Through its peer-to-peer topology, MAPLE exhibits great scalability: the higher the mobile host density, the more queries can be answered by peers.*

## 1 Introduction

Ubiquitous and untethered information access is stimulated by the growing capabilities of mobile devices. Location-based services are especially of interest to many users. An example query might be to “find the nearest gas station from my current location.” The combination of Global Positioning Systems (GPS), wireless communication technology (e.g., 802.11x), and peer-to-peer overlays (P2P) offers an exciting environment and opens unique opportunities to provide location based services. By leveraging ad-hoc networks, information can be shared in a P2P manner among mobile clients to answer spatial searches (e.g., nearest neighbor queries). Importantly, in cases where access to the remote database servers is not always guaranteed – such as during a natural disaster – P2P sharing can provide a robust alternative where fault-resilience is naturally built into the design.

Recently we designed a novel technique that utilizes previous query results cached in the local storage of mobile peers to compute nearest neighbor query results of mobile hosts [3]. We introduced a search algorithm termed *Sharing-based Nearest Neighbor* ($\mathrm{SN}^2$) to locally verify whether points of interest (POI) received from peers provide a complete, partial or irrelevant answer to a submitted location-based nearest neighbor query. If the POI search can be fully satisfied from the peers, no access to any remote spatial database servers is necessary, resulting in a decreased workload. Under some circumstances only a partial result can be algorithmically verified (details are provided later) and in that case, the query must be forwarded to a remote spatial database server. However, the server load can be diminished by utilizing the partial result to constrain the search space.

In this simulation system we present MAPLE (Mobile scAlable Peer-to-peer nearest neighbor query model for Location-based sErvices), a prototype of the design in [3] that implements a sharing-based NN query model in conjunction with a road network environment. In particular, the MAPLE system exhibits the following distinguishing characteristics:

- **Sharing-based nearest neighbor query execution.** The MAPLE system performs and visualizes a novel sharing-based nearest neighbor ($\mathrm{SN}^2$) query algorithm among peers in a step-by-step manner.
- **Scalability.** The MAPLE system leverages P2P sharing to achieve scalability in terms of the number of peers and to reduce the access frequency to remote spatial database servers. A higher density of peers improves its efficiency.
- **Realistic movement on road network.** The movement of hosts in the MAPLE system is constrained to real world road networks. Mobile hosts in MAPLE autonomously proceed on road networks and the velocity of the movement is determined by the speed limit of underlying road segments.

---

*This research has been funded in part by NSF grants EEC-9529152 (IMSC ERC), CMS-0219463 (ITR), and equipment gifts from the Intel Corporation, Hewlett-Packard, Sun Microsystems and Raptor Networks Technology.

## 2 The MAPLE Infrastructure

Figure 1 illustrates the infrastructure of MAPLE. We are focusing on mobile peers, such as cars, that are instrumented with a GPS to provide continuous location information. Furthermore, we assume that two tiers of wireless connections are available on future automobiles. Traditional, cellular-based networks (such as utilized by the OnStar service) allow medium range connections to base-stations that interface with the wired Internet infrastructure. A second type of short-range networks allows ad-hoc connections with neighboring mobile clients. Technologies that enable short range communication include, for example, IEEE 802.11x. Benefiting from the power capacities of vehicles, we assume that each mobile host has a significant transmission range and virtually unlimited lifetime. The architecture can also support hand-held mobile devices. However, then power consumption becomes an additional parameter which we are not currently considering.

In our design, previous query results can be cached in the local storage of mobile peers. Such peers move on road networks and autonomously launch sharing-based NN queries or exchange cached query results with adjacent peers. The $SN^2$ algorithm is then applied to verify whether data items received from neighboring peers provide a complete, partial, or irrelevant answer to the posed query. If only partial or irrelevant data items are collected, the query is forwarded to a spatial database server [4]. The complete query result is then also cached in the local memory.

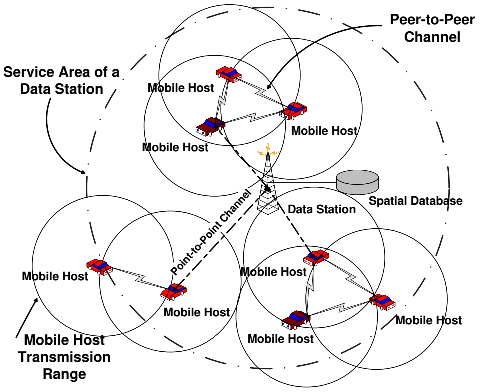


**Figure 1. The MAPLE infrastructure.**

## 3 The MAPLE Components

In this section we describe the modular implementation of the MAPLE system and the simulation parameter set such as the points of interest and mobile host density.

### 3.1 Modules

The MAPLE system consists of three components: (1) the *multiple peer simulation module*, (2) the *server module*, and (3) the *sharing-based nearest neighbor query visualization module*. Sharing-based NN queries are demonstrated by visualizing the interactions and data exchange processes among the peers on the road network.

- The **multiple peer simulation module** concurrently models a predefined number of mobile hosts. It implements all the functionality of a single mobile host and provides the communication facilities among peers and from peers to remote spatial database servers.
- The **server module** is responsible for storing points of interest indexed by an R-tree structure [2]. It performs NN queries from peers with pruning bounds and records the I/O load and access frequency of the spatial database server.
- The **sharing-based nearest neighbor query visualization module** provides a rendering of the verification process of a sharing-based NN query in a step-by-step manner. Users can arbitrarily select a mobile host and launch a location-based NN query within the simulation region.

### 3.2 Road Network Generation

We generated the underlying road network from the TIGER/LINE street vector data set available from the U.S. Census Bureau. The current MAPLE system stores the road network of several Southern California counties. The road segments are differentiated into several road classes, such as freeways, primary highways, secondary and connecting roads, and rural roads. Road segments of different road classes are associated with different driving speed limits. Mobile peers monitor the speed limit on the road that they are currently traveling on and adjust their velocity accordingly. One of the challenges when integrating road segments into a complete road network is to isolate paths that cross and determine if they indeed represent intersections. For example, freeways generally project many intersections in two-dimensional space, however, many of them are over-passes or bridges. Our solution is to detect intersection points with the help of their endpoint coordinates. In addition, differing road classes allow us to distinguish over-passes from intersections.

### 3.3 Interest Objects and Mobile Peers

MAPLE models the density of POIs (currently gas stations and restaurants) in the Greater Los Angeles area via data available from two online sites: GasPriceWatch.com[1] and CNN/Money. MAPLE also imports vehicle statistics of the Greater Los Angeles area from the Federal Statistics web site to initialize the mobile host density. However, users

---

[1] http://www.gaspricewatch.com

are not limited to utilizing these density presets. Many parameters – the density of interest objects and mobile peers among them – can be changed via the multiple peer simulation module.

### 3.4 Sharing-based Nearest Neighbor Queries and Pruning Bounds

Within the system infrastructure shown in Figure 1, a mobile host $Q$ collects NN data from peers to harvest these existing results for completing its own $k$ nearest neighbor ($k$NN) search. If data items collected from peers only ensure a partial answer, then queries need to be forwarded to the spatial database server to retrieve the complete result. However, the partial result can be used to bound and hence speed up the server search process.

There are two approaches to process NN information obtained from peers. The single peer NN verification process, also called $k\text{NN}_{single}$, attempts to verify whether a point of interest $n_i$ obtained from a peer is a valid (i.e., top $k$) nearest neighbor of a mobile host $Q$. To this end we utilize the spatial relationship between mobile hosts and their POIs as follows.

**Theorem 3.1** *Let $Q$ and $P_1$ be two mobile hosts, and let $P_1$ have $k$ nearest neighbors, $n_1, n_2, \ldots, n_k$, which are sorted in ascending order according to their distance to $P_1$. For any nearest neighbor $n_i$ of $P_1$, if Dist(Q, $n_i$) + $\delta \leq$ Dist($P_1$, $n_k$) then $n_i$ is one of the top $k$ nearest neighbors of $Q$.*

In Theorem 3.1, Dist($Q$, $n_i$) is the Euclidean distance between $Q$ and $n_i$, $\delta$ is the Euclidean distance between $Q$ and $P_1$, and Dist($P_1$, $n_k$) is the Euclidean distance between $P_1$ and its cached farthest nearest neighbor $n_k$. An illustration of Theorem 3.1 is shown in Figure 2. The nearest neighbor $n_2$ of mobile host $P_1$, which is a peer of mobile host $Q$, can be verified as the nearest neighbor of $Q$.

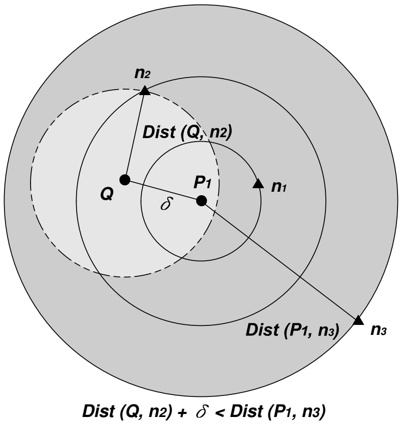


**Figure 2. Verification of a point of interest.**

Under some conditions the $k\text{NN}_{single}$ method may not be able to verify all $k$ nearest neighbors. Therefore, we extend the verification process to include results from multiple peers simultaneously. The $k\text{NN}_{multiple}$ method combines the area of all the peers, each bounded by the outermost NN circle, into a *certain region* $R_c$. The $k\text{NN}_{multiple}$ verification technique is executed based on $R_c$ similarly to $k\text{NN}_{single}$. Theorem 3.2 provides the rules for verifying nearest neighbors with multiple peers.

**Theorem 3.2** *If the nearest neighbor data set $NN_P$ is composed of data from $j$ peers, the certain region $R_c$ can be represented as:*

$$R_c = P_{1-area} \cup P_{2-area} \cup \cdots \cup P_{j-area}.$$

*For any interest object $n_i$ in $R_c$, the distance between $Q$ and $n_i$ is used as a radius to draw a circle $C_{ni}$. If $C_{ni}$ is fully covered by $R_c$, then $n_i$ is a valid NN of $Q$.*

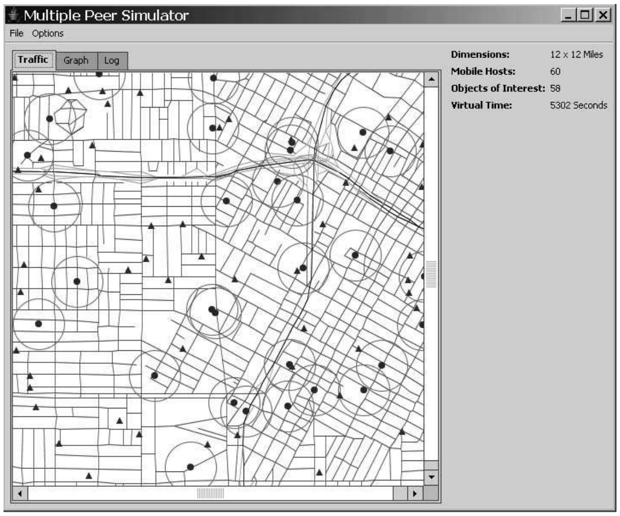
There will be cases when neither $k\text{NN}_{single}$ nor $k\text{NN}_{multiple}$ can fulfill a $k$NN query. Hence a set with some unverified elements is returned. If the response time is critical, a user may agree to accept such a $k$NN data set, where the objects are not guaranteed to be the top $k$ nearest neighbors. Otherwise the $k$NN query must be forwarded to a spatial database server. The partial results can be used to bound – and hence speed up – the server search process [3].

### 3.5 System Demonstration

MAPLE is implemented in Java and Figure 3 shows its multiple peer simulator interface. The left window frame shows the simulated service region of MAPLE visualizing all mobile hosts and POIs. Mobile hosts move on the road network autonomously [1] while observing the speed limit constraints. On the right pane, the simulator displays the configuration parameters of current simulation, such as the service region dimensions, and the number of mobile hosts. Users are able to select (via mouse click) any mobile host to launch a $k$NN query. This action activates the sharing-based NN query algorithm visualization interface shown in Figure 4. Each POI retrieved from peers within the communication range is first verified in a step-by-step manner using the $k\text{NN}_{single}$ method. In the case that the $k\text{NN}_{single}$ method cannot verify all $k$ nearest neighbors, the simulator automatically launches the $k\text{NN}_{multiple}$ method, illustrated in Figure 5. Figure 6 depicts the server interface of MAPLE. The server module executes the spatial queries received from peers. It also records the page access frequency within the spatial database server when performing spatial queries.

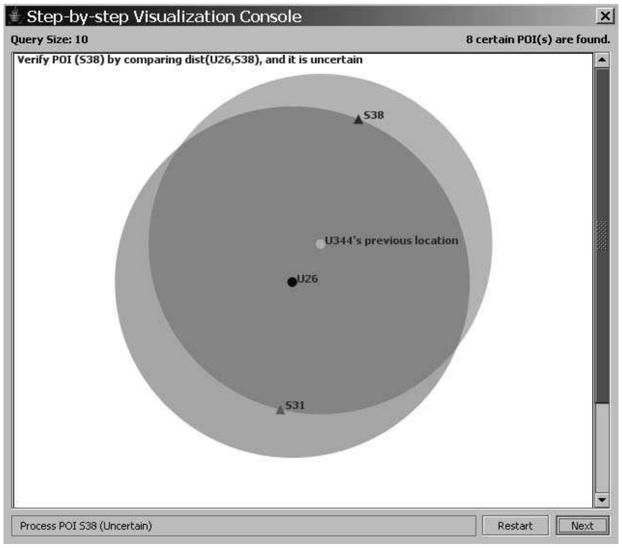
## 4 Conclusions

We have described MAPLE, a system to aid in the study of scalable peer-to-peer data sharing for location-based ser-



Dimensions: 12 x 12 Miles
Mobile Hosts: 60
Objects of Interest: 58
Virtual Time: 5302 Seconds

**Figure 3. The multiple peer simulator interface of MAPLE.**



Query Size: 10

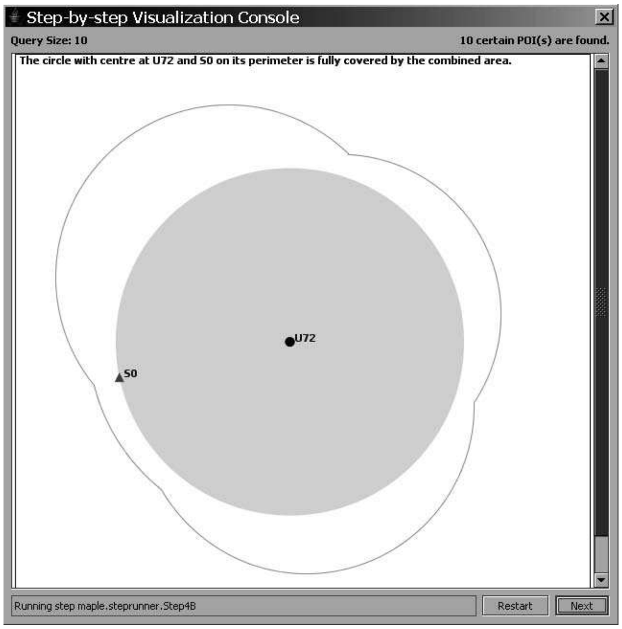
8 certain POI(s) are found.

Verify POI (S38) by comparing dist(U26,S38), and it is uncertain

Process POI S38 (Uncertain)

**Figure 4. The sharing-based NN query algorithm visualization interface of MAPLE illustrating a single peer verification example.**

vices in mobile environments. We implemented localized verification algorithms to ascertain whether data items received from peers provide a relevant answer to posed queries. We also implemented a road network with realistic speed limits to constrain the movement of peers in the mobile environment. MAPLE is interactive and provides step-by-step visualizations to understand peer communication and verification procedures. The objective of MAPLE is to provide a platform for the evaluation of our ongoing research into the novel design of P2P sharing techniques for location-based services. MAPLE demonstrates the excellent scalability and



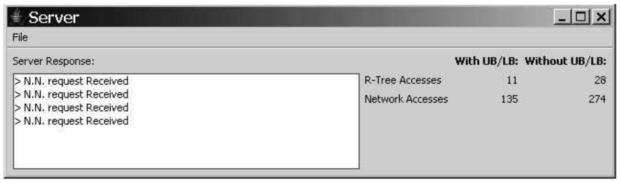
Query Size: 10

10 certain POI(s) are found.

The circle with centre at U72 and S0 on its perimeter is fully covered by the combined area.

Running step maple.steprunner.Step4B

**Figure 5. The sharing-based NN query algorithm visualization interface of MAPLE illustrating a multiple peer verification example.**



Server Response:

> N.N. request Received
> N.N. request Received
> N.N. request Received
> N.N. request Received

| | With UB/LB: | Without UB/LB: |
|---|---|---|
| R-Tree Accesses | 11 | 28 |
| Network Accesses | 135 | 274 |

**Figure 6. The server interface of MAPLE.**

effectiveness of our current algorithm in high density mobile environments.

## References

[1] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. In *Proceedings of the* $4^{th}$ *ACM/IEEE MobiCom*, pages 85–97, 1998.

[2] A. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 47–57, Boston, Massachusetts, June 18-21, 1984.

[3] W.-S. Ku, R. Zimmermann, and C.-N. Wan. Location-based Spatial Queries with Data Sharing in Mobile Environment. Technical Report USC-CS-TR05-843, University of Southern California, 2005.

[4] R. Zimmermann, W.-S. Ku, and W.-C. Chu. Efficient Query Routing in Distributed Spatial Databases. In *12th ACM International Workshop on Geographic Information Systems.*, pages 176–183, 2004.