Mobile Local Search via P2P Databases[1]

Yan Luo, Ouri Wolfson, and Bo Xu

Department of Computer Science, University of Illinois, Chicago, IL, USA

{yluo, wolfson, boxu}@cs.uic.edu

Abstract – We examine the problem of mobile local search via P2P databases. We argue that P2P databases can lead to a new form of mass communication that may revolutionize local search. We introduce our approach, MOBI-DIK, which is a platform-independent search engine based on mobile P2P databases.

1. Introduction

Mobile local search is a procedure in which a mobile user searches for local resources, i.e., resources that are in geographic proximity to the user (e.g., a person with certain expertise in a convention hall, a ride-share opportunity, a taxi-cab, a parking slot, etc.). In this paper we examine mobile local search by a P2P paradigm. In this paradigm, a database is stored in the peers[2] (PDA's, cell phones, vehicles, sensors, etc.) of a Mobile Ad-hoc NETwork (MANET) (Figure 1.1). These peers communicate with each other via broadband (typically tens of Mbps) but short-range wireless technologies such as IEEE 802.11, Bluetooth, Zigbee, or Ultra Wide Band (UWB). On each mobile peer there is a local database that stores and manages a collection of reports, where each report is a data item that describes a resource; often the resource is dynamic, i.e. it is temporarily available (e.g. a parking slot, or person of interest close by). A report is sensed, or entered by the user of a mobile peer, or otherwise obtained by the peer. For the purpose of search, a report is analogous to a web page in Google.

All the local databases maintained by the mobile peers form a mobile P2P database (Figure 1.2). The peers communicate reports and queries (a query is analogous to a keyword-search in Google terminology) to neighbors directly, and the reports and queries propagate by transitive multi-hop transmissions; thus a search of the database is not limited to the immediate neighbors. In contrast to traditional information management systems, an infrastructure and a central server may not exist, and the information hosts are autonomous and mobile.

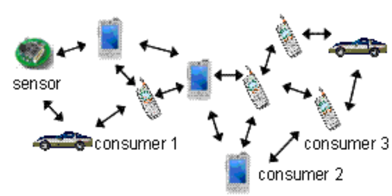

Figure 1.1. Mobile ad hoc solution to local search

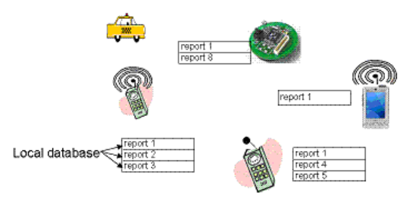

Figure 1.2. A mobile P2P database

The characteristics of mobile P2P databases include (i) dynamic network topology, (ii) memory/energy/bandwidth limitations, and (iii) lack of global coordination. The objective of our MOBI-DIK (MOBIle DIscovery of local resource Knowledge) project is to develop a mobile P2P search engine in this environment. By the nature of mobile P2P interactions, the search is limited to local information. The MOBI-DIK search engine is platform-independent, and it can lead to new forms of mass communication that may revolutionize search. Another way of looking at a mobile P2P search engine is as a bridge between the virtual world (reports and queries) and the physical world (resources). The purpose of this paper is to motivate and introduce MOBI-DIK. The rest of paper is organized as follows. Section 2 discusses several applications that use mobile local search. Section 3 discusses and compares the mobile P2P and the centralized architectures. Section 4 presents our MOBI-DIK approach to mobile local search. Section 5 concludes the paper and addresses the future research directions. Appendices A and B compare MOBI-DIK with existing mobile P2P algorithms.

[1] Research supported by US NSF Grants DGE-0549489, 0513736, 0326284.

[2] Peers are equal in terms of the communication and the data management protocol they use. But they are not necessarily the same device type nor have the same capabilities.

2. Mobile Local Search Applications

Social Networks

In a large professional, political, or social gathering, MOBI-DIK is useful to automatically facilitate a face-to-face meeting based on matching profiles. For example, in a professional gathering, MOBI-DIK enables attendees to specify queries (interest profiles) and resource descriptions (expertise) to facilitate conversations, when mutual interest is detected. This opportunistic matchmaking can greatly enhance the value of networking events allowing users to connect with targeted, interested parties without a priori knowledge of their name, title, phone number, or other personal information. A face-to-face meeting can be setup by including in the resource description the identification information of the resource (person), such as cell-phone number, name, screen name, picture, physical description, etc. This information may be used together with the (possibly imprecise) location to help set up the face-to-face meeting. Thus, the individual's profile that is stored in MOBI-DIK will serve as a "wearable web-site". Similarly, MOBI-DIK can facilitate face-to-face meetings in singles matchmaking.

Emergency Response, Homeland Security, and the Military

MOBI-DIK offers the capability to extend decision-making and coordination capability. This finds applications in emergency environments, to support the more than eight million first responders[3] in U.S. homeland security. Consider workers in disaster areas, soldiers and military personnel operating in environments where the wireless fixed infrastructure is significantly degraded or non-existent. A mobile P2P database lets them automatically propagate messages, pictures, or resource information to other workers, based on matching profiles, security, and attribute values rather than node-id. As mobile users involved in an emergency response naturally cluster around the location of interest, a self-forming, high-bandwidth network that allows secure point-to-point or point-to-multipoint communication without the need of potentially compromised infrastructure could be of great benefit. Another potential application of MOBI-DIK is shipping container monitoring and inspection, in which sensors mounted on neighbouring containers can communicate and transitively relay alerts to remote checkpoints. MOBI-DIK can be also used for search/dissemination of real-time intelligence, surveillance and reconnaissance battlefield data.

Mobile E-commerce

Consider mobile devices in a mall obtaining merchants' sale and inventory information via short-range wireless broadcast or embedded RFID readers. The mobile P2P database will enable a customer (whose cell phone is MOBI-DIK enabled) that enters a mall to locate a desired product at the best price. It will also enable merchants to automatically and frequently change prices and offer real-time sales notices and promotions for items.

Airport Applications

From the point of view of commerce, airports have stores and kiosks where merchandise as in a mall. Imagine arriving at a large airport and realizing you do not have the computer cord you need for your presentation. MOBI-DIK will enable a user to search for the needed product - just like in a mall. MOBI-DIK can also be used by airport personnel to coordinate their activities. This is especially important when there is a communication failure due an emergency that degrades the infrastructure. Another potential opportunity is the search and dissemination of real-time information regarding flight changes, delays, queue length, parking information, special security alerts and procedures, and baggage information.

Urban Transportation Safety and Efficiency

Inefficiencies in the transportation system result in excessive environmental pollution, fuel consumption, risk to public safety, and congestion. The average annual delay due to traffic congestion has climbed over 300% in the past two decades, going from 7 hours spent stuck in traffic per person per year in 1982 to 26 hours in 2001 [1]. MOBI-DIK software can improve safety and mobility by enabling travelers to cooperate intelligently and automatically. A vehicle will be able to automatically and transitively communicate to trailing vehicles its "slow speed" message when it encounters an accident, congestion, or dangerous road surface conditions. This will allow other drivers to make decisions such as finding alternative roads. Also, early warning messages may allow a following vehicle to anticipate sudden braking, or a malfunctioning brake light, and thus prevent pile-ups in some situations. Similarly, other resource information, such as ridesharing opportunities, transfer protection (transfer bus requested to wait for passengers), available parking slots and taxi cabs, will be propagated transitively, improving the efficiency of transportation.

---

[3] First responders are the personnel of organizations and agencies such as emergency medical services; fire, rescue, and hazardous material response teams; security and law enforcement agencies; relief organizations.

NASA Applications

NASA's exploration of planetary surfaces will require dissemination of local information by surface-based assets, including base stations, astronauts, habitats, rovers, and robots. The local resources of interest to mobile users are often only available during a limited period of time and these resources themselves may be mobile. For example, a malfunctioning robot is available only until it completely loses capability. Similarly, the current location of an astronaut, and the current temperature in a particular geographic area, are temporarily valid or available resources. MOBI-DIK provides surface-based mobile assets with an efficient information dissemination and search tool.

3. Substituting or Augmenting a Centralized Architecture

Consider mobile users that search for local resources. Assuming that the information about the existence and location of such a resource resides on a server, a communication infrastructure is necessary to access the server. Such an infrastructure may not be available in military/combat situations, disaster recovery, in a commercial flight, etc. Even if the infrastructure and a server are both available, a user may not be willing to pay the dollar-cost that is usually involved in accessing the server through the cellular infrastructure. Furthermore, cellular bandwidth is limited (e.g. 130 character text messages). In other words, a client-server approach may have accessibility problems.

Currently, Google and local.com provide static local information (e.g. the location of a restaurant, pharmacy, etc.), but not dynamic information such as the location of a taxi cab, a nearby person of interest, or an available parking slot. These dynamic resources are temporary in nature, and thus require timely, real-time update rates. Such rates are unlikely to be provided for the country or the world by a centralized server farm, a la Google. Thus, dynamic local resources may require local servers, each dedicated to a limited geographic area. However, for many areas such a local server may not exist due to lack of a profitable business model, and if it exists it may be unavailable (such servers are unlikely to have the reliability of global sites such as Google). Furthermore, the data on the server may be unavailable due to propagation delays (think of sudden-brake information that needs to be propagated to a server and from there to the trailing vehicles), or due to device limitations (e.g. a cab customer's cell-phone may have Bluetooth but not internet access to update the server), or due to the fact that updates from mobile devices may involve a communication cost that nobody is willing to pay, or due to the fact that the local server (e.g. of Starbucks) may accept only updates from certain users or certain applications but not others. In short, a client-(local)-server may have both accessibility and availability problems.

We propose to substitute or augment the client-(local)-server approach by a MANET database. The MANET database is distributed among the mobile peers. Communication in the MANET is free since it uses the unlicensed spectrum, and larger in bandwidth than the cellular infrastructure, thus can provide media rich information, such as maps, menus, and even video. A mobile user may search the MANET database only, or combine it with a client-server search.

From the historical point of view, the currently prevalent centralized architecture was designed at a time when the mobile devices did not possess the capability to make them intelligent. And therefore, in such an architecture, the intelligence of mobile communication lies in the fixed network infrastructure, and not in the devices. However, nowadays, mobile devices have hundreds-of-MHz processors, and tens of megabytes of random access memory. So another way of looking at this proposal is as a shift of paradigm towards intelligent devices.

4. MOBI-DIK: Our Approach

There are two main paradigms for answering queries in mobile P2P databases, one is report pulling and the other is report pushing.

Report pulling means that a mobile peer makes an explicit request for the report it is interested in receiving, and the whole network is flooded with queries, the interested report will be pulled from the mobile peers that have them. Report pulling is widely used in resource discovery, such as route discovery in mobile ad hoc networks and file discovery by query flooding in wired P2P networks like Gnutella. Flooding in a wireless network is in fact relatively efficient as compared to wired networks because of the wireless multicast advantage.

Another possible approach for data dissemination is report pushing; reports are flooded, and consumed by peers whose query is answered by the received reports. So far there exist mechanisms to broadcast information in the complete network, or in a specific geographic area (geocast), apart from to any one specific mobile node (unicast/mobile ad-hoc routing) or any one arbitrary node (anycast). The report pushing paradigm can be further divided into stateful methods and stateless methods. Most stateful methods are topology-based, i.e. they impose a structure on the network, and maintain states of data dissemination. PStree, which organizes the peers as a tree, is an example of topology based methods [2].

Report pulling and stateful approaches suffer when the network topology is dynamic and disconnected. Thus we propose a store-and-forward algorithm, MOBI-DIK, based on the following ideas and results:

1. In MOBI-DIK the growing-local-database problem of store-and-forward algorithms is addressed by prioritization; each mobile peer prioritizes the reports in order to accommodate them in limited power, bandwidth, and memory. The priority of a report depends on its size (the larger the report, the more resources it consumes), demand (how many peers are querying it), and supply (how many peers already have it). The demand of a report is estimated by sampling of the queries that it satisfies, but sampling does not work for estimating the supply. Thus,

   a. We examined estimating the supply of a report based on various indicators such as the age of the report and the number of times it is received by a mobile peer. We found that no single indicator is a good predictor of supply in all environments. For example, in some environments the intuition that the age of the data-item is a good predictor of novelty is correct. However, in an environment where many new peers are entering the system, the number of times a report is received by a mobile peer is a much better indicator of supply.

   b. We developed and implemented an algorithm called MALENA that combines various indicators to estimate the current supply [3]. The combination uses a machine learning system that is trained from previously received reports, and automatically learns the best indicator for the current environment.

   c. We compared MALENA with three approaches to ranking of reports [3]. In the first two approaches, ranking is based on a single individual supply indicator that is found to be an optimal indicator in an environment that disfavors another. The third approach is PeopleNet [4]. The comparison is conducted by simulations using real Bluetooth traces collected at a major conference event. The comparison shows that in each individual environment MALENA approaches or outperforms the best algorithm for that environment and outperforms the inferior one by up to 5 times. This is important since the best algorithm depends on the global environment, and the global environmental parameters change and are usually not known to a mobile peer.

   d. Simulations show that, with each mobile peer storing 1000 reports and allocating 10% of its power to MALENA, the performance of MALENA reaches 56% of the ideal benchmark (where a central server is employed, and it instantaneously broadcasts each new report to all the mobile peers). 30% of the reports received by MALENA are received within 10 minutes. We also calculated the CPU consumption of MALENA. On a typical PDA, each execution of the MALENA computation takes less than half a millisecond.

2. An additional issue arising in a Store-and-forward algorithm such as MOBI-DIK is how many reports to communicate in each transmission. If too many, excessive collisions arise, and if too few, then the search capability suffers. We developed an analytical model that computes the throughput of a transmission in an 802.11 ad hoc network. The throughput is computed based on collision factors such as the transmission size, the transmission frequency, the density of mobile peers, and so on (see [5]). Using this analytical model we proposed a method by which a mobile peer dynamically adjusts the P2P transmission size depending on the period of time between two P2P transmissions of a peer. The objective is to optimize the bandwidth and energy utilization; optimization occurs when the number of reliably received reports delivered per unit of energy is maximized. Simulations show that by dynamically adjusting the transmission size, the performance of the mobile P2P search is improved by up to an order of magnitude.

3. In MOBI-DIK a peer propagates reports when a new neighbor is encountered or when new reports are received (old reports to new neighbors, or new reports to old neighbors), thus adapting to both low and high mobility environments.

4. We compared MOBI-DIK with the three existing mobile P2P algorithms, namely RANDI [6], PeopleNet [4], and 7DS [7]. MOBI-DIK outperforms PeopleNet and 7DS by an order of magnitude, and is up to two times better than RANDI. The advantage of MOBI-DIK in a practical application, namely parking slot discovery, which was demonstrated in [8].

5. We studied how the availability of a fixed infrastructure (e.g. the internet) can be exploited to augment the P2P data dissemination (absence of a central server is assumed). The general idea is that when a match (i.e. a query and a report that match each other) occurs at a broker, the broker sends the matching report to the query originator via the infrastructure. Since the cellular transmission is, per byte, 16 times more energy-costly than P2P transmission, it is not even clear that backchannel communication is beneficial. However, we determined that it is.

6. We studied the application of MOBI-DIK in the dissemination of real-time traffic information. In this application, the real-time traffic information is produced by and serviced to vehicles throughout a road network. This application complements the existing real-time traffic information dissemination methods which tend to cover only selected highways where speed sensors are deployed. We compared MOBI-DIK with Grassroots [17], a flooding based mobile P2P traffic information dissemination algorithm. The results show that MOBI-DIK outperforms Grassroots when the vehicle density is sparse or when the available bandwidth is small. In some cases MOBI-DIK outperforms Grassroots by 50%. These results demonstrate the benefit of store-and-forward, information prioritization, and bandwidth adaptation.

7. We implemented a simplified version of the data management layer on the Pocket PC/WinCE platform. It allows two PDA's to conduct Bluetooth and Wifi search sessions. Using two STTR commercialization grants we are preparing for a trial at the Dallas/Fort-Worth International airport.

5. Research Directions for Mobile P2P Local Search

Further work is necessary on data query processing for mobile P2P search applications. Work on sensor databases (e.g. Tinydb [9], POS [13]) addresses data-models and languages for sensors, but considers query processing in an environment of static peers. Cartel [10] addresses the translation of these abstractions to an environment in which cars transfer collected data to a central database via fixed access points. Work on MANET protocols deals mainly with routing and multicasting. In this landscape there is a gap, namely general query-processing in a MANET's; such processing probably needs to be cognizant of many issues related to peer-mobility. For example, existing mobile P2P query processing methods deal with simple queries, e.g. selections; each query is satisfied by one or more reports. However, in many application classes one may be interested in more sophisticated queries, e.g. aggregation. For instance, in mobile electronic commerce a user may be interested in the minimum gas price within the next 30 miles on the highway. Processing of such P2P queries may present interesting optimization opportunities.
After information about a mobile resource is found, localization is often critical for finding the physical resource. However, current self-localization techniques are insufficient. For example, GPS is not available indoors and the accuracy of GPS is not reliable. Thus, furthering the state of the art on localization (see e.g., [11]) is important for mobile P2P search.

As discussed above, mobile P2P databases do not guarantee answer completeness. In this sense, the integration with an available infrastructure such as the internet or a cellular network may improve performance significantly. This integration has two aspects. First, using the communication infrastructure in order to process queries more efficiently; and second, using data on the fixed network in order to provide better and more answers to a query. The seamless integration of mobile P2P databases and infrastructure databases introduces important research challenges.

Other important research directions include: incentives for broker participation in query processing (see [12]), and transactions/atomicity/recovery issues in databases distributed over mobile peers (virtual currency must be paid in an atomic fashion).

Of course, work on efficient resource utilization in mobile peers, and coping with sparse networks and dynamic topologies (see e.g., [14]) is still very important for mobile P2P search.

Reference

[1] The 2003 Urban Mobility Report, available at http://tti.tamu.edu/product/catalog/reports/

[2] Y. Huang and H. Garcia-Molina, "Publish/subscribe in a Mobile Environment", *MobiDE*, 2001.

[3] O. Wolfson and B. Xu, "Flooding by Machine Learning in Mobile Ad Hoc Networks", accepted, to appear in *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*.

[4] M. Motani, V. Srinivasan, and P. Nuggehalli, "PeopleNet: Engineering a Wireless Virtual Social Network", *Proc. of ACM MobiCom*, 2005.

[5] O. Wolfson, B. Xu, H. Yin, and H. Cao, "Search-and-Discover in Mobile P2P Network Databases", *ICDCS*, July 2006.

[6] O. Wolfson and B. Xu, "Mobile Peer-to-peer Data Dissemination with Resource Constraints", *MDM*, 2007.

[7] M. Papadopouli and H. Schulzrinne, "Effects of Power Conservation, Wireless coverage and Cooperation on Data Dissemination Among Mobile Devices", *Proc. of ACM MobiHoc*, 2001.

[8] B. Xu, O. Wolfson, C. Naiman, N. Rishe, and R. Tanner, "A Feasibility Study on Disseminating Spatio-temporal Information via Vehicular Ad-hoc Networks", *V2VCOM*, 2007.

[9] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an Acquisitional Query Processing System for Sensor Networks", ACM Trans. on Database System, 30(1):122–173, 2005.

[10] B. Hull, et al., "CarTel: A Distributed Mobile Sensor Computing System", In Proc. of 4th Intl. Conf. on Embedded Networked Sensor Systems, pp. 125–138, Nov 2006.

[11] M. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca, F. Potter, I. Smith, A. Varshavsky. "Practical Metropolitan-Scale Positioning for GSM Phones", Ubicomp 2006.

[12] B. Xu, O. Wolfson, and N. Rishe, "Benefit and Pricing of Spatio-temporal Information in Mobile Peer-to-Peer Networks", Proc. of Hawaii Intl. Conf. on System Sciences (HICSS-39), Jan 2006.

[13] L. Cox, M. Castro, and A. Rowstron, "POS: Practical Order Statistics for Wireless Sensor Networks", *26th IEEE ICDCS*, Lisbon, Portugal, July 2006.

[14] W. Zhao, M. Ammar, E. Zegura. "Controlling the Mobility of Multiple Data Transport Ferries in a Delay Tolerant Network", *INFOCOM* 2005.

[15] R. Barr. An efficient, unifying approach to simulation using virtual machines. PhD thesis, Cornell University, 2004.

[16] D. Choffnes and F. Bustamante. An Integrated Mobility and Traffic Model for Vehicular Wireless Networks. *Second ACM Workshop on Vehicular Ad Hoc Networks* (VANET 2005), Cologne, Germany, September 2, 2005.

[17] S. Goel, T. Imielinski, K. Ozbay. Ascertaining viability of WiFi based vehicle-to-vehicle network for traffic information dissemination. The 7th *International IEEE Conference on Intelligent Transportation Systems*, 2004, pp. 1086-1091.