

The MOBI-DIK Approach to Searching in Mobile Ad Hoc Network Databases¹

Yan Luo, Ouri Wolfson, and Bo Xu
Department of Computer Science
University of Illinois at Chicago

ABSTRACT

In this chapter, we introduce the mobile ad-hoc network (MANET) database by discussing its definition, historical background and scientific fundamentals. Existing related projects are presented and classified into two main categories, pedestrian and vehicular projects based on their target users. Two main paradigms (i.e., report pulling and report pushing) for answering queries in MANET databases are discussed in details. Then we present the MOBIDIK approach to searching in MANET databases and compare it with alternatives. Finally, the key applications and the future research directions are addressed.

INTRODUCTION: MANET DATABASES

A Mobile Ad-Hoc Network (MANET) database is a database that is stored in the peers of a MANET. The network is composed by a finite set of mobile peers that communicate with each other via short range wireless protocols, such as IEEE 802.11, Bluetooth, Zigbee, or Ultra Wide Band (UWB). These protocols provide broadband (typically tens of Mbps) but short-range (typically 10-100 meters) wireless communication. On each mobile peer there is a local database that stores and manages a collection of data items, or reports. A report is a set of values sensed or entered by the user at a particular time, or otherwise obtained by a mobile peer. Often a report de-

¹ Research supported by NSF Grants DGE-0549489, and IIS- 0513736, 0847680, and 0326284.

scribes a physical resource such as an available parking slot. All the local databases maintained by the mobile peers form the MANET database. The peers communicate reports and queries to neighbors directly, and the reports and queries propagate by transitive multi-hop transmissions. Figure 1 below illustrates the definition.

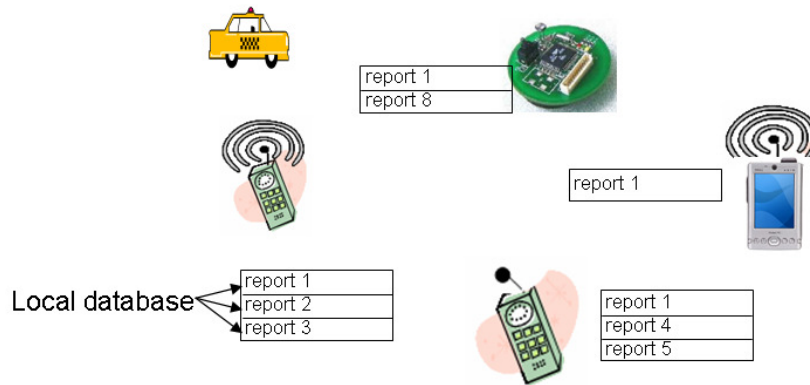


Figure 1: A MANET database

MANET databases enable matchmaking or resource discovery services in many application domains, including social networks, transportation, mobile electronic commerce, emergency response, and homeland security.

Communication is often restricted by bandwidth and power constraints on the mobile peers. Furthermore, often reports need to be stored and later forwarded, thus memory constraints on the mobile devices constitute a problem as well. Thus, careful and efficient utilization of scarce peer resources (specifically bandwidth, power, and memory) are an important challenge for MANET databases.

HISTORICAL BACKGROUND

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 in the cloud, 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 have the properly equipped device (e.g. the device may have Bluetooth but not Wi-Fi), or may not be willing to pay the dollar-cost that is usually involved in accessing the server through the cellular infrastructure. In other words, a cloud-computing search may have accessibility and cost 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 current exact 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 that cannot be provided by web crawling.

In cloud-computing search, the searches issued by a user may be tracked by a search engine and used to find the identity of the user. The association of searches with the user identity may generate serious privacy concerns, a lesson given by the recent AOL incident (see [25][26]). A concern that is particular to conducting local search in the cloud is that the location of the user is potentially disclosed when the user provides his/her area of geographical interest to the search engine. Furthermore, some information, such as the status of home appliances, is simply not worth putting in the cloud.

Cloud-computing search can be complemented by a MANET database search. The MANET database is distributed among local peers, thus the MANET search does not depend on availability of an infrastructure. This also makes the MANET database more amenable to host the information that is otherwise not worth putting in the cloud. Furthermore, communication among mobile peers is free since it uses the unlicensed spectrum, thus the cost concern is addressed. The response time of MANET database search is higher because the database of neighboring peers can be accessed faster than the web is crawled. Finally, MANET database search addresses the privacy concern because the P2P communication can be totally anonymous,

and searches as well as the database are distributed among autonomous entities.

Currently, there are quite a few experimental projects in MANET databases. These can be roughly classified into pedestrians and vehicular projects. Vehicular projects deal with high mobility and high communication topology change-rates, whereas pedestrians projects have a strong concern with power issues. The following are several active experimental MANET database projects for pedestrians and vehicles:

Pedestrians Projects	Features
7DS – Columbia University ([14]) http://www.cs.unc.edu/~maria/7ds/	Focuses on accessing web pages in environments where only some peers have access to the fixed infrastructure.
iClouds – Darmstadt University http://iclouds.tk.informatik.tu-darmstadt.de/	Focuses on the provision of incentives to brokers (intermediaries) to participate in the mobile P2P database.
MoGATU –University of Maryland, Baltimore County http://mogatu.umbc.edu/	Focuses on the processing of complex data management operations, such as joins, in a collaborative fashion.
PeopleNet – National University of Singapore ([13]) http://www.ece.nus.edu.sg/research/projects/abstract.asp?Prj=101	Proposes the concept of information Bazaars, each of which specializes in a particular type of information; reports and queries are propagated to the appropriate bazaar by the fixed infrastructure.
MoB – University of Wisconsin and Cambridge University	Focuses on incentives and the sharing among peers of virtual

http://www.cs.wisc.edu/~suman/projects/agora/	information resources such as bandwidth.
Mobi-Dik – University of Illinois at Chicago http://www.cs.uic.edu/~wolfson/html/p2p.html	Focuses on information representing physical resources, and proposes stateless algorithms for query processing, with particular concerns for power, bandwidth, and memory constraints.

Vehicular Projects	Features
CarTALK 2000 – A European project http://www.cartalk2000.net/	Develops a co-operative driver assistance system based upon inter-vehicle communication and mobile P2P databases via self-organizing vehicular ad-hoc networks.
FleetNet – Internet on the Road Project http://www.ccrle.nec.de/Projects/fleetnet.htm	Develops a wireless multi-hop ad hoc network for intervehicle communication to improve the driver's and passenger's safety and comfort. A data dissemination method called "contention-based forwarding" (CBF) is proposed in which the next hop in the forwarding process is selected through a distributed contention mechanism based on the current positions of neighbors.
VII – Vehicle Infrastructure Integration, a US DOT project http://www.its.dot.gov/vii/	The objective of the VII project is to deploy advanced vehicle-to-vehicle (using the mobile P2P paradigm) and vehicle-to-infrastructure communications

	that could keep vehicles from leaving the road and enhance their safe movement through intersections.
Grassroots – Rutgers University ([22]) http://paul.rutgers.edu/~gsamir/dataspace/grassroots.html	Develops an environment in which each vehicle contributes a small piece of traffic information to the network based on the P2P paradigm, and each vehicle aggregates pieces of the information into a useful picture of the local traffic information.

SCIENTIFIC FUNDAMENTALS

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

Report pulling means that a mobile peer issues a query which is flooded in the whole network, and the answer-reports will be pulled from the mobile peers that have them (see e.g. [1]). 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 broadcast advantage, but there are also disadvantages which will be explained below.

Another possible approach for data dissemination is report pushing. Report pushing is the dual problem of report pulling; reports are flooded, and consumed by peers whose query is answered by 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). 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 of links in the network, and maintain states of data dissemination. PStree [2], which organizes the peers as a tree, is an example of topology based methods.

Another group of stateful methods is cluster- or hierarchy-based method, such as [3], in which moving peers are grouped into some clusters or hierarchies and the cluster heads are randomly selected. Reports are disseminated through the network in a cluster or hierarchy manner, which means that reports are first disseminated to every cluster head, and each cluster head then broadcasts the reports to the member peers in its group. Although cluster- or hierarchy-based methods can minimize the energy dissipation in moving peers, these methods will fail or cost more energy in highly mobile environments since they have to maintain a hierarchy structure and frequently re-select cluster heads.

Another stateful paradigm consists of Location-based methods (see [4]). In location-based methods, each moving peer knows the location of itself and its neighbors through some localization techniques, such as GPS or Atomic Multilateration (see [4]).

The simplest location-based data dissemination is Greedy Forwarding, in which each moving peer transmits a report to a neighbor that is closer to the destination than itself. However, Greedy Forwarding can fail in some cases, such as when a report is stuck in local minima, which means that the report stays in a mobile peer whose neighbors are all further from the destination. Therefore, some recovery strategies are proposed, such as GPSR (Greedy Perimeter Stateless Routing [5]). Other location-based methods, such as GAF (Geographic Adaptive Fidelity [6]) and GEAR (Geographical and Energy Aware Routing [7]), take advantage of knowledge about both location and energy to disseminate information and resources more efficiently.

In stateless methods, the most basic and simplest one is flooding-based method, such as [8]. In flooding-based methods, mobile peers simply propagate received reports to all neighboring mobile peers until the destination or maximum a hop is reached. Each report is

propagated as soon as is received. Flooding-based methods have many advantages, such as no state maintenance, no route discovery, and easy deployment. However, they inherently cannot overcome several problems, such as implosion, overlap, and resource blindness. Implosion refers to the waste of resources taking place when a node forwards a message to a neighbor although the latter may have already received it from another source. Overlap occurs when two nodes read the same product record or coupon, and thus push into the network the same information. Resource blindness denotes the inability of the protocol to adapt the node's behaviour to its current availability of resources, mainly power [9]. Therefore, other stateless methods are proposed, such as gossiping-based methods and negotiation-based methods.

Gossiping-based methods, such as [10], improve flooding by transmitting received reports to a subset of randomly selected neighbors; another option is to have some neighbours simply drop the report. For example, the neighbors that are not themselves interested in the report drop it. The advantages of gossiping-based methods include reducing the implosion and lowering the system overhead. However, dissemination, and thus performance, is reduced compared to pure flooding.

Negotiation-based methods solve the implosion and overlap problem by transmitting first the id's of reports; the reports themselves are transmitted only when requested (see [11]). Thus, some extra data transmission is involved, which costs more memory, bandwidth, and energy. In addition, in negotiation-based methods, moving peers have to generate meta-data or a signature for every report so that negotiation can be carried out, which will increase the system overhead and decrease the efficiency.

Another important stateless paradigm for data dissemination in MANET databases is store-and-forward. In contrast to flooding, store-and-forward does not propagate reports as soon as they are received; rather they are stored and rebroadcast later. This obviously introduces storage and bandwidth problems, if too many reports need to be saved and rebroadcast at the same time. To address these,

methods such as [12] rank all the reports in a peer's database in terms of their relevance (or expected utility), and then the reports are communicated and saved in the order of their relevance. Or, the reports requested and communicated are the ones with the relevance above a certain threshold. The notion of relevance quantifies the importance or the expected utility of a report to a peer at a particular time and at a particular location. Other store-and-forward methods include PeopleNet [13] and 7DS [14].

In summary, the paradigms for data dissemination in MANET databases are summarized in figure 2 below.

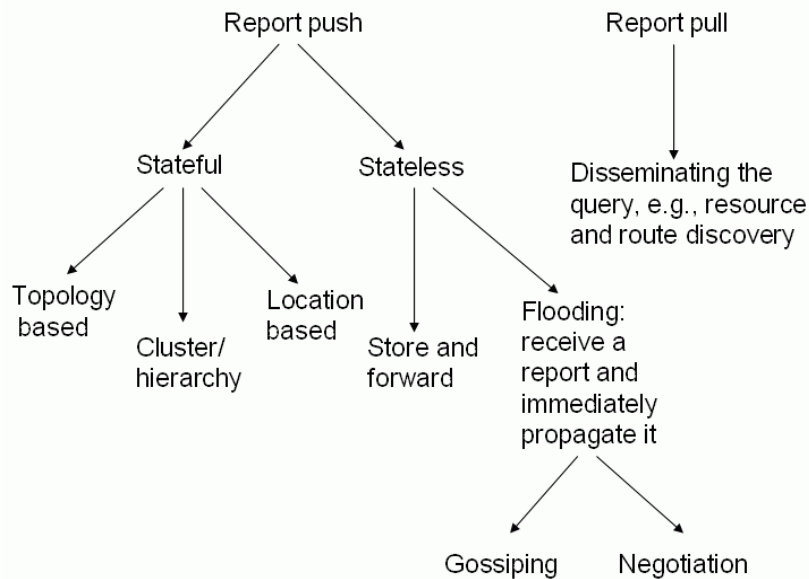


Figure 2: Query answering methods in MANET databases

THE MOBI-DIK APPROACH

Report pulling and stateful approaches suffer when the network topology is dynamic, disconnected, and/or sparse. 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, because supply increases continuously as the report is being disseminated. 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 [19]. The combination uses a machine learning system that is trained from previously received reports, and automatically learns the best indicator for the current environment. Deleted: [19]
 - c. We compared MALENA with three approaches to ranking of reports [19]. 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 [13]. 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 Deleted: [19]

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 (see [12]). The throughput is computed based on collision factors such as the transmission size, the transmission frequency, the density of mobile peers, and so on. 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 [20], PeopleNet [13], and 7DS [14]. MOBI-DIK outperforms PeopleNet and 7DS by an order of magnitude, and is up to two times better than RANDI. More details are provided in Appendix A. The advantage of MOBI-DIK in a practical application, namely parking slot discovery, was demonstrated in [21].
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. More details are provided in Appendix A.
6. We demonstrated the MOBI-DIK algorithm in the context of specific queries and reports. The queries are K-nearest-neighbor queries, and the reports are the current locations of mobile sensors. Thus, the MOBI-DIK algorithm is specialized to process this specific spatial query in a data-to-query fashion. Similarly, it can be specialized for in-network processing of other types of queries, e.g. spatial window queries. We compare the MOBI-DIK continuous-KNN performance with that of the in-network KNN algorithm given in [27], called DIKNN. It processes queries in a query-to-data fashion. The results show that MOBI-DIK is up to 50 times more accurate than DIKNN when the P2P network is sparse, but DIKNN is more accurate when the network is dense. More details are provided in appendix B.
7. 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 disseminated 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 [22], 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. More details are provided in Appendix C.

KEY APPLICATIONS

MANET databases provide mobile users a search engine for transient and highly dynamic information in a local geospatial environment. MANET databases employ a unified model for both the cellular infrastructure and the mobile ad hoc environments. When the infrastructure is available, it can be augmented by the MANET database approach.

Consider a MANET database platform, i.e. a set of software services for data management in a MANET environment; it is similar to a regular Database Management System, but geared to mobile P2P interactions. Such a platform will enable quick building of matchmaking or resource discovery services in many application domains, including social networks, emergency response and homeland security, the military, airport applications, mobile e-commerce, and transportation.

1. Social Networks

In a large professional, political, or social gathering, MANET databases are useful to automatically facilitate a face-to-face meeting based on matching profiles. For example, in a professional gathering, MANET databases enable attendees to specify queries (interest profiles) and resource descriptions (expertise) to facilitate face-to-face meetings, when mutual interest is detected. Thus, the individual's profile that is stored in MANET databases will serve as a "wearable web-site". Similarly, MANET databases can facilitate face-to-face meetings for singles matchmaking.

2. Emergency Response, Homeland Security, and the Military

MANET databases offer the capability to extend decision-making and coordination capability. Consider workers in disaster areas, soldiers and military personnel operating in environments where the wireless fixed infrastructure is significantly degraded or non-existent. As mobile users involved in an emergency response naturally cluster around the location of interest, a self-forming, high-bandwidth network that allows database search without the need of potentially compromised infrastructure could be of great benefit. For instance, the search could specify a picture of a wanted person.

3. Airport Applications

A potential opportunity that will benefit both the consumer and the airport operations is the dissemination and querying of real-time information regarding flight changes, delays, queue length, parking information, special security alerts and procedures, and baggage information. This can augment the present audio announcements that often cannot be heard in nearby restaurants, stores, or restrooms, and augment the limited number of displays.

4. Mobile E-commerce

Consider short-range wireless broadcast and mobile P2P dissemination of a merchant's sale and inventory information. It will enable a customer (whose cell phone is query-capable) that enters a mall to locate a desired product at the best price. When a significant percentage of people have mobile devices that can query retail data, merchants will be motivated to provide inventory/sale/coupons information electronically to nearby potential customers. The information will be disseminated and queried in a P2P fashion (in, say, a mall or airport) by the MANET database software.

5. Transportation Safety and Efficiency

MANET databases 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), will be propagated transitively, improving the efficiency of the transportation system.

FUTURE RESEARCH DIRECTIONS

Further work is necessary on data models for mobile P2P search applications. Work on sensor databases (e.g. Tinydb [15]) addresses data-models and languages for sensors, but considers query processing in an environment of static peers (see e.g., POS [16]). Cartel [17] 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 MANET's; such processing 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, existing (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 is important for mobile P2P search.

As discussed above, MANET 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 MANET databases and infrastructure databases introduces important research challenges.

Other important research directions include: incentives for broker participation in query processing (see [18]), and transactions/atomicity/recovery issues in databases distributed over mobile peers (virtual currency must be transferred from one peer to another in an atomic fashion, otherwise may be lost).

Of course, work on efficient resource utilization in mobile peers, and coping with sparse networks and dynamic topologies is still very important for mobile P2P search.

REFERENCES

- [1] S. M. Das, H. Pucha, and Y. C. Hu, "Ekta: An Efficient DHT Substrate for Distributed Applications in Mobile Ad hoc Networks", in *Proc. of the 6th IEEE Workshop on Mobile Computing Systems and Applications*, English Lake District, UK, 12 2004.
- [2] Y. Huang and H.G. Molina, "Publish/subscribe in a Mobile Environment", *MobiDE*, 2001.
- [3] A. Visvanathan, J. H. Youn, and J. Deogun, "Hierarchical Data Dissemination Scheme for Large Scale Sensor Networks," in *IEEE International Conference on Communications (ICC'05)*, pp. 3030-3036, May 2005.
- [4] M. Mauve, A. Widmer, and H. Hartenstein, "A Survey on Position-Based Routing in Mobile Ad-Hoc Networks," *IEEE Network*, vol. 15, no. 6, pp. 30-39, 2001.

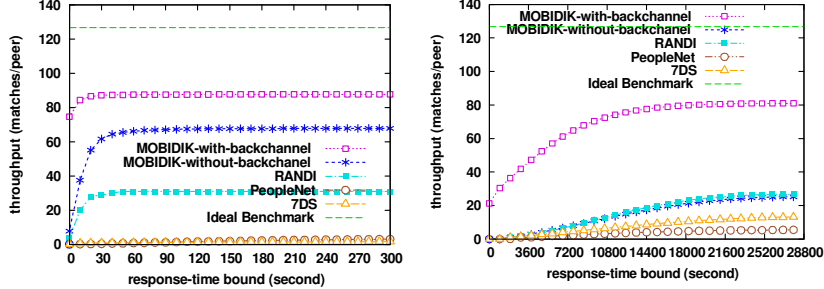
- [5] B. Karp and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Sensor Networks," in *The 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'00)*, pp. 243-254, Aug 2000.
- [6] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed Energy Conservation for Ad hoc Routing," in *The ACM International Conference on Mobile Computing and Networking*, pp. 70-84, Rome, Italy, July 2001.
- [7] Y. Yu, R. Govindan, and D. Estrin., "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks," *Technical Report UCLA/CSD-TR-01-0023, UCLA*, May 2001.
- [8] R. Oliveira, L. Bernardo, and P. Pinto, "Flooding Techniques for Resource Discovery on High Mobility MANETs", *Workshop on Wireless Ad-hoc Networks*, 2005.
- [9] A. A. Papadopoulos and J. A. McCann, "Towards the Design of an Energy-Efficient, Location-Aware Routing Protocol for Mobile, Ad-hoc Sensor Networks", in *Proc. of the 15th Int. Workshop on Database and Expert Systems Applications (DEXA)*, 2004.
- [10] A. Datta, S. Quarteroni, and K. Aberer, "Autonomous Gossiping: A Self-Organizing Epidemic Algorithm For Selective Information Dissemination in Wireless Mobile Ad-Hoc Networks," in *The International Conference on Semantics of a Networked World*, 2004.
- [11] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-Based Protocols for Disseminating Information in Wireless Sensor Networks," *Wireless Networks*, vol. 8, pp. 169-185, 2002.
- [12] O. Wolfson, B. Xu, H.B. Yin, and H. Cao, "Search-and-Discover in Mobile P2P Network Databases", in *Proc. of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06)*, Lisbon, Portugal, 2006
- [13] M. Motani, V. Srinivasan, and P. Nuggehalli, "PeopleNet: Engineering a Wireless Virtual Social Network," *International Conference Mobile Computing and Networking (MobiCom'05)*, Aug 2005.

- [14] M. Papadopouli and H. Schulzrinne, "Design and Implementation of a P2P Data Dissemination and Prefetching Tool for Mobile Users", *First NY Metro Area Networking Workshop*, IBM TJ Watson Research Center, Hawthorne, NY, 2001.
- [15] 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.
- [16] L. Cox, M. Castro, and A. Rowstron, "POS: Practical Order Statistics for Wireless Sensor Networks", *26th IEEE ICDCS*, Lisbon, Portugal, July 2006.
- [17] 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.
- [18] B. Xu, O. Wolfson, and N. Rische, "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.
- [19] B. Xu, O. Wolfson, and C. Naiman, "Flooding by Machine Learning in Mobile Ad Hoc Networks", accepted, to appear in *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*.
- [20] O. Wolfson and B. Xu, "Mobile Peer-to-peer Data Dissemination with Resource Constraints", *MDM*, 2007.
- [21] B. Xu, O. Wolfson, C. Naiman, N. Rische, and R. Tanner, "A Feasibility Study on Disseminating Spatio-temporal Information via Vehicular Ad-hoc Networks", *V2VCOM*, 2007.
- [22] S. Goel, T. Imielinski, K. Ozbay, "Ascertaining Viability of WiFi based Vehicle-To-Vehicle Network for Traffic Information Dissemination," *The 7th Intel. IEEE Conf. on Intelligent Transportation Systems*, 2004.
- [23] R. Barr, "An Efficient, Unifying Approach to Simulation Using Virtual Machines", PhD thesis, Cornell University, 2004.
- [24] D. Choffnes and F. Bustamante, "An Integrated Mobility and Traffic Model for Vehicular Wireless Networks", *VANET*, 2005.

- [25] S. Hansell. AOL Removes Search Data On Vast Group Of Web Users, *New York Times*, August 8, 2008. <http://query.nytimes.com/gst/fullpage.html?res=9504E5D81E3FF93BA3575BC0A9609C8B63>
- [26] M. Barbaro and T. Zeller. A Face Is Exposed for AOL Searcher No. 4417749, *New York Times*, August 9, 2008. http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=1&scp=11&sq=google%20search%20privacy%20woman&st=cse
- [27] Wu, S.-H., et al: Diknn: an itinerary-based knn query processing algorithm for mobile sensor networks. *ICDE'07*.
- [28] B. Karp and H. Kung. Gpsr: Greedy perimeter stateless routing for wireless networks. In *Proc. of MOBICOM*, 2000.

Appendix A. Comparison of MOBI-DIK with RANDI, PeopleNet, and 7DS

From the algorithmic point of view, the main differences of 7DS and PeopleNet from MOBI-DIK are: 1) they have no energy/bandwidth management for determining the transmission size; 2) the broker function is much more simplistic (no ranking); 3) 7DS does not have a good strategy to determine when to communicate. The main difference of RANDI from MOBI-DIK is that it does not take supply into account when ranking reports. The four algorithms are implemented in SWANS (Scalable Wireless Ad-hoc Network Simulator) built at Cornell ([23]). We augmented SWANS with a feature that tracks the energy consumed for computation and for 802.11 communication (including transmitting, receiving, and listening). In order to put the performance of the algorithms in perspective, we also define an ideal benchmark offline algorithm. In the ideal benchmark algorithm, when a report is produced, it is instantaneously disseminated to all the mobile peers currently in the system.



(a) On average each peer has 20 neighbors

(b) On average each peer has 0.1 neighbor

Figure A.1: Comparison among MOBI-DIK, RANDI, PeopleNet, and 7DS. 802.11 bandwidth=2Mbps, transmission range=100 meters, 15% of battery energy allocated to a mobile P2P algorithm, 1 report produced every 10 seconds, mean of report size=1500 bytes, query size = 100 bytes, mean of reports database size=100K bytes

Figure A.1 shows some simulation results. The performance measure, response-time bounded throughput, was developed by us and it integrates two traditional evaluation metrics for data access, namely throughput and response-time. Specifically, the response-time bounded throughput is the average number of distinct matches received by a mobile peer within a certain time limit c . c is called the response-time bound. The following observations are made from the simulation results:

1. With at least 15% of the battery energy allocated to each mobile peer, a mobile P2P database performs reasonably close to an ideal centralized database.
2. MOBI-DIK outperforms PeopleNet and 7DS by an order of magnitude.
3. Supply-and-demand ranking doubles the performance of a mobile P2P database (see Figure A.1(a)).
4. The backchannel improves the performance of MOBI-DIK by nearly 30% even though it is more energy costly.
5. The backchannel is particularly beneficial when the peer density is low.

Appendix B: Query-to-data and data-to-query in P2P KNN-Query Processing

In this appendix we demonstrate the MOBI-DIK algorithm in the context of K -nearest-neighbors (KNN) queries, where the reports are the current locations of mobile sensors. Thus, the MOBI-DIK algorithm is specialized to process this specific spatial query. We compare the MOBI-DIK continuous-KNN performance with that of the in-network KNN algorithm given in [27] (called DIKNN). This section is organized as follows. In §B.1 we discuss the instantiation of MOBI-DIK for KNN query processing. In §B.2 we compare MOBI-DIK and the DIKNN algorithm ([27]) which is designed especially for KNN query processing in mobile sensor networks.

B.1 KNN Query Processing in MOBI-DIK

We consider a system where there is a single static sink. The sink is continuously interested in knowing the k mobile peers that are closest to a static query point q . Thus the sink and every other peer have a query Q which contains the coordinates of q . Q is known to all the peers before the KNN query processing and therefore it is not transmitted during a P2P interaction. A report is produced by each mobile peer O every one second. The report, referred to as the *location report*, contains the current timestamp, the current location of O , and the peer-id of O . We refer to the location included in a report R as the *home location* of R .

For the computation of the demand, $Q(R)$, i.e., the degree to which a report R satisfies the KNN query Q is negatively correlated to the distance between the home location of R and the query point q , and is [0-1] normalized. For example, if the distance between the home location of R and q is 1.4 mile, and the maximum distance between q and the home location of a report in the peer's database is 2 miles, then the demand for R is $(1-1.4/2=0.3)$.

The reports propagate via P2P to the sink, which is one of the peers. The sink answers the KNN queries based on the latest location of each peer maintained in the sink's reports database.

B.2 Comparison Between Query-to-data by DIKNN ([27]), and Data-to-query by MOBI-DIK

The DIKNN Algorithm. In the literature there is no algorithm for in-network processing of continuous KNN queries. Thus we use an instantaneous algorithm, DIKNN, and execute it repeatedly. In DIKNN, the query issued by the sink is geographically routed from the sink to the nearest neighbor of the query point p by GPSR ([28]). The query is then disseminated to all the peers in a circular search-

ing area centered at p . The size of the searching area is determined based on the density of the network, such that the k nearest neighbors are inside the searching area with a high probability. At the end of the dissemination, the aggregated query response is routed back to the sink.

Simulation Method. The mobile P2P algorithms are implemented in SWANS (Scalable Wireless Ad-hoc Network Simulator) built at Cornell. 100 peers move within a $L \times L$ square area according to the *random way-point* mobility model with mean speed 1 mile/hour and mean pause time 180 seconds. In this setup, the average number of peers D within a circle with radius 100 meters (i.e., the transmission range), is

$$D = 100 \times \frac{\pi \times 100 \times 100}{L^2}$$

We refer to D as the *peer density*. D is a system parameter. It has values 20 and 1, representing high and low connectivities. In other words, even though MOBI-DIK is geared towards a sparse connectivity environment, we evaluate it in dense ones as well. L is chosen such that the peer density equals to the value of D . The whole simulation runs for 8 simulated hours.

The sink is located at the center of the northwest quadrant of the simulated square area. The query point is located at the center of the southeast quadrant. The query is issued every one second. For MOBI-DIK, this means that KNN's are computed at the sink using the local database every 1 second. The local database has the last known location of each peer, but some of these locations may be outdated, thus the computed KNN may not be accurate. The accuracy of each KNN query is measured by the fraction of the correctly returned KNN's.

The size of each report is 24 bytes, including 16 bytes for the coordinates of the location, 4 bytes for the timestamp, and 4 bytes for the peer-id. The size limit of each reports database is randomly chosen from [1200, 3600] bytes. Thus the average number of members in a reports database is $2400K/24=100$. All the simulation parameters and their values are listed in Table B.1.

For DIKNN, we do not simulate its operation. Instead, we compute the upper bound of its query accuracy. Assume that the geographic routing is reliable and instantaneous, and the searching area contains all the K -nearest neighbors, thus the query is correctly delivered to them. Then with DIKNN, a correct KNN is returned to the sink if and only if there exists a path between this KNN and the sink at the time when the query is issued. For each query, we compute the fraction of the correct KNN's for which there are paths between them and the sink, and take this fraction to be the accuracy of DIKNN.

The above simulation method favors MOBI-DIK in the sense that MOBI-DIK is designed for continuous queries and DIKNN is designed for instantaneous queries. On the other hand, it favors DIKNN in the following senses. First, DIKNN is optimized for KNN query processing whereas MOBI-DIK is a general purpose query processing algorithm. Second, MOBI-DIK is compared with the upper bound accuracy of DIKNN.

TABLE B.1. SIMULATION PARAMETERS AND THEIR VALUES

Parameter		Unit	Values
Number of concurrent peers			100
Total number of peers			100
Transmission range	r	meter	100
Peer density	D	Number of peers in a circle with radius r	1, 2, 3, 4, 5, 6, 7
Data transmission speed		bits/second	2M
Battery allocation fraction	F		1
Report production rate	P	reports/second/peer	1
Report size range		byte	24
Query size		byte	16
Reports database size range		byte	[1200, 3600]
Demand database size		byte	16
Average motion speed		miles/hour	1
Average pause time		second	180
Mobility model			Random way-point
Simulation run time		hours	8

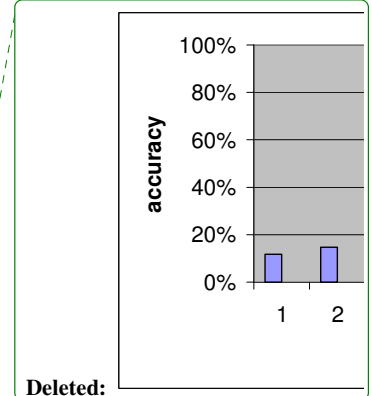
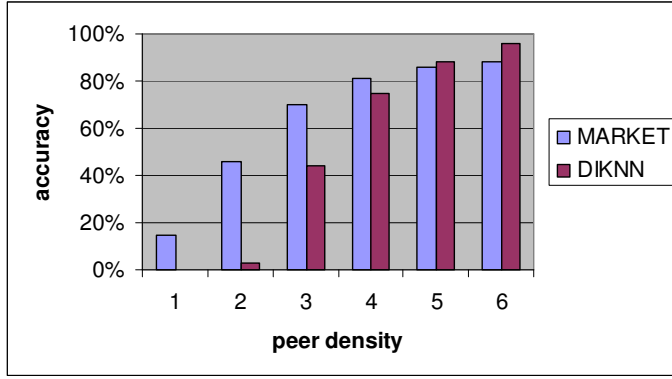
Deleted: , 7

Simulation results. Figure B.1 shows the accuracy of MOBI-DIK and DIKNN when k is 10, for different peer densities. Figures B.2 and B.3 show the accuracy when peer density is 2 and 1 respectively, for different values of k . From these figures it can be seen that the accuracy of MOBI-DIK is by far higher than that of DIKNN when the peer density is low. When the peer density is below 2, the DIKNN algorithm completely fails, with 0% accuracy. In this case MOBI-DIK manages to provide up to 30% accuracy (see Figure B.3). When the peer density is 2, MOBI-DIK outperforms DIKNN by an order of magnitude, regardless of the value of k . In this case the accuracy of MOBI-DIK reaches 60% for $k=20$ (see Figure B.2). The reason for the poor performance of DIKNN in a sparse network is that, with very high probability there does not exist a path between the sink and a KNN. When the sink and the KNN are disconnected, the query does not reach the KNN and the KNN is not collected in the query response. MOBI-DIK, on the other hand, does not need a contemporaneous path to collect results. This demonstrates the benefit of store-and-forward in a sparse network.

Deleted: 5

Deleted: 5

When the peer density is higher than 5 (see Figure B.1), DIKNN starts to outperform MOBI-DIK.



Deleted:

Deleted: ies 0.5 and

Figure B.1. Accuracy as a function of peer density with $k=10$. The accuracy is zero for DIKNN for peer density 1.

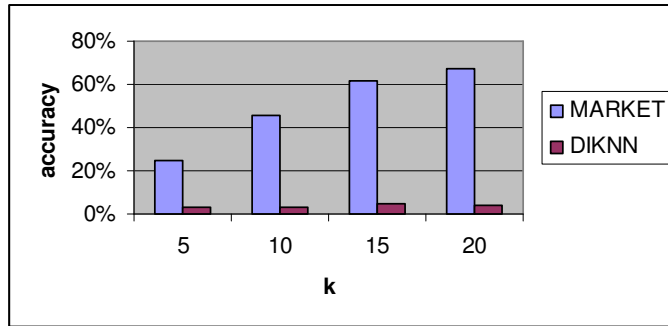


Figure B.2. Accuracy as a function of k with peer density = 2

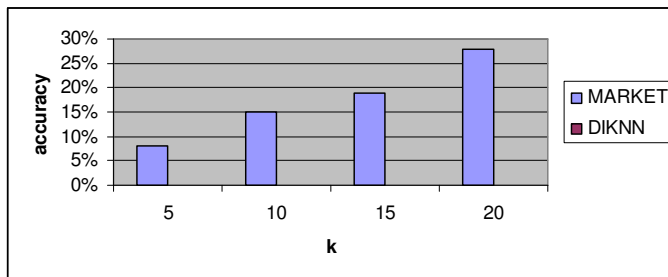


Figure B.3. Accuracy as a function of k with peer density = 1. The accuracy is zero for DIKNN for all the k values.

Appendix C. Comparison of MOBI-DIK with Grassroots in Dissemination of Real-time Traffic Information

The application scenario is as follows. A set of vehicles move on a road network. Each vehicle m has a *digital map* of the road network. For each road segment s in the road network, the digital map stores the estimated travel time of s . Each vehicle m is equipped with a GPS receiver. Every time m travels through a road segment s and reaches the end of it, m produces a report that includes the travel time experienced by m on s . The report is disseminated using a mobile P2P algorithm (MOBI-DIK or Grassroots). Upon receiving a report, m updates the estimated travel time of the corresponding road segment in its own digital map.

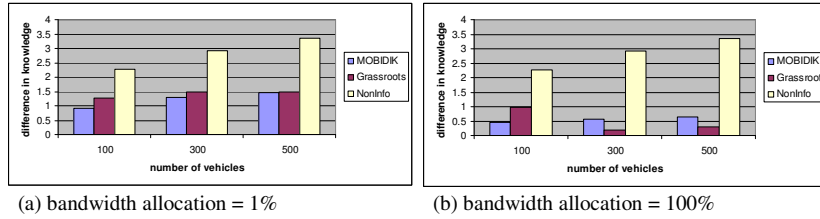
We propose a novel metric, called the difference in knowledge (DIK), as the performance measure. The DIK measures the difference between the actual traffic condition on a road segment and that known to a vehicle. Formally, let s_1, s_2, \dots, s_n be all the road segments in the road network. Let p be the location of a vehicle O at t . Let $\mathbb{T}(s_k)$ be the actual travel time of a road segment s_k . Let $T(O, s_k)$ be the travel time of s_k maintained at O 's digital map. The difference in knowledge of O at t , denoted $DIK(O, t)$ is

$$DIK(O, t) = \sum_{k=1}^n \left(\frac{1}{g_k} |T(s_k) - T(O, s_k)| \right) \quad (4.1)$$

where g_k is the free-flow travel time along the shortest-distance path from p to the middle point of s_k . Intuitively, the difference in knowledge of O is the weighted sum of the difference in knowledge of O between the true map and its local digital map for each road segment. The farther away the road segment is from O , the lower its weight. Specifically, the weight of a road segment s_k is $1/g_k$.

To put the results in perspective we evaluated MOBI-DIK and Grassroots against the case in which reports are not disseminated at all. By comparing with this case, called NonInfo, we evaluate the benefit of reports dissemination. The comparison uses the

STRAW/SWANS simulation test-bed ([24]). The STRAW system simulates vehicle traffic mobility on a 3.2km \times 2.2km region of downtown Chicago.



(a) bandwidth allocation = 1%

(b) bandwidth allocation = 100%

Figure C.1: Comparison among MOBI-DIK, Grassroots, and NonInfo. 802.11 bandwidth=2Mbps, transmission range=250 meters,

Deleted: B

Figure C.1 shows some simulation results. From the figure it can be seen that both MOBI-DIK and Grassroots significantly outperform NonInfo, by up to an order of magnitude. Furthermore, when the vehicle density is small, MOBI-DIK outperforms Grassroots; this demonstrates the advantage of store-and-forward. When a small fraction of 802.11 bandwidth is allocated to traffic information dissemination (the rest may be reserved for emergent dissemination of safety-related messages such as a malfunctioning brake or a deployment of an airbag), MOB-DIK also outperforms Grassroots; this demonstrates the benefit of information prioritization and bandwidth adaptation. Grassroots outperforms MOBI-DIK when the bandwidth allocation and vehicle density are high, and therefore we are defining a system that adaptively transitions between the two algorithms depending on these parameters.

Deleted: B