Decentralized Spatial Computing in Urban Environments

Patrick Laube[1], Matt Duckham[1], Mike Worboys[2], and Tony Joyce[3]

[1]Geomatics Department, The University of Melbourne, Australia

[2]National Center for Geographic Information and Analysis, University of Maine, Orono, ME

[3]Ordnance Survey of Great Britain, Southampton, UK

This chapter presents the concept of *decentralized spatial computing* (DeSC) as a way to embed dynamic spatial data capture and processing capabilities within our built urban environment. The chapter illustrates the potential of DeSC for safeguarding privacy in a dynamic location-based services scenario: Mobile service users protect their potentially sensitive location by the use of a decentralized query algorithms, solely collaborating with peers close by and thereby excluding the privacy bottleneck of an omniscient global service provider. In an extensive set of consecutive experiments several decentralized query algorithms were tested, trading the level of privacy for the quality of service. The use of a real world test bed, — a small part of Ordnance Survey's OS MasterMap® Integrated Transport Network™ Layer for Southampton — underlines the experiments' validity. The chapter concludes with a research and development agenda for DeSC in the urban context.

Keywords: decentralized spatial computing, ambient spatial intelligence, mobile wireless sensor networks, privacy, location-based services.

1 Introduction

In a world of dynamic, networked, and data-rich computing, the days of spatial data processing in a monolithic geographic information system are numbered. Ubiquitous, embedded, and highly distributed computing has led to the emergence of *pervasive computing*, *ubiquitous computing*, *ambient intelligence* (AmI), and 'everyware' (Greenfield, 2006). Similarly, spatial computing systems are becoming highly dynamic, multi-party networks, in which mobile human users interact in dynamic networks, and spatially distributed autonomous computing nodes form the 'cyber-infrastructure' of urban environments. These spatial applications of 'everywhere', termed here *ambient spatial intelligence* (AmSI), are becoming increasingly important to the natural and built environments of the future.

The goal of AmSI is to embed spatial data capture and processing capabilities within the environment itself, for example using *wireless sensor networks* (WSN). Traditional forms of centralized, client/server spatial data processing are hardly capable of enabling AmSI. Amongst other issues, centralized approaches to the highly dynamic, real-time nature of AmSI data sources lead to information and communication overload, unmasking the unscalable nature of conventional GIS and spatial database architectures. Omniscient centralized databases furthermore present a potential privacy breach. Hence, AmSI rather requires adopting decentralized architectures, where spatially distributed but collaborating computing nodes autonomously take on responsibility for responding to spatial queries with no centralized control (Laube & Duckham, in press). For example, with respect to privacy protection, such decentralized architectures allow private and sensitive spatial data to be collected at different sites, and analyzed in a decentralized way without collating and storing personal data in a centralized GIS or spatial database.

The emergence of such *decentralized spatial computing* (DeSC) is so far most evident in the area of environmental monitoring, where wireless geosensor networks are challenging conventional ways of centralized modeling and detecting change (Duckham, Nittel, & Worboys, 2005; Worboys & Duckham, 2006) The same paradigm shift is now also reaching the urban context, where spatial information processing provides the backbone of a range of application domains, including location-based services (LBS), traffic management, and facilities management. Hence, this chapter explores the notion of DeSC for embedding AmSI in urban environments.

AmSI has also raised clear privacy concerns, due to its potential for real time monitoring and rapid integration of personal and sensitive location information (Dobson & Fisher, 2003). The same technological advancement that allows AmSI to pervade our urban environments also leads to the availability of finer and finer granularities of location information about users of LBS. With ever finer spatio-temporal granularity, however, location information becomes a *quasi-identifier*, allowing re-identification of previously anonymized information (Bettini, Wang, & Jajodia, 2005). Hence, when considering the future of AmSI, balancing the quality of provided services and the level of privacy sacrificed for those services, becomes a key challenge.

In this chapter, we examine the potential for using DeSC techniques to deliver high quality, dynamic location-based services within an AmSI environment, at the same time as protecting the privacy of mobile individuals accessing those services. Our approach is explicitly dynamic: instead of protecting individual location fixes, we aim to protect 'trajectory privacy' — the degree to which aggregated knowledge of an individual's location over time can be used to invade that person's location privacy. In our decentralized approach, mobile individuals query their spatial neighbors for responses to spatial queries. Since the neighborhood of a mobile individual is constantly changing, the likelihood that a single hostile agent ever collects enough information to seriously threaten any particular individual's location privacy is decreased. In other words, mobile individuals 'smear' their location information across spacetime in order to protect their privacy. For many common spatial and LBS queries (for example, *k*-nearest neighbors) this strategy can still enable relatively high quality of service because it can exploit the spatial structure of decentralized knowledge, where mobile agents that are closer in space are more likely to possess information relevant to one another.

In conclusion, the major contributions of this chapter are:

- a discussion of the potential of DeSC in an urban context;
- a case study applying and testing the concept of DeSC for safeguarding privacy in a LBS application; and
- a road map for further DeSC research and development in the urban context.

2 Related Work

Recent work in at least three distinct areas is particularly relevant to our work. This section provides an overview and synthesis of these three related topics: current paradigm shifts in process-oriented spatiotemporal modeling of urban environments; the fundamentals of decentralized spatial computing; and location privacy protection for LBS.

2.1 The city in flux

Urban environments are highly dynamic. On various spatial and temporal scales, urban events and processes range from urban sprawl through migration over commuter traffic flows to pedestrian movement in a mall. Or as Worboys and Hornsby (2004, p. 327) state figuratively, 'processes of urban growth and decline, migration, and expansion, constitute the city in flux'. LBS and intelligent travel assistance are just two examples that illustrate the potential of AmSI and DeSC in urban environments (Dillenburg, Wolfson, & Nelson, 2002). For example, Winter and Nittel (2006) have shown for shared ride trip planning that decentralization is not only scalable but can deliver near-optimal solutions with local knowledge only.

Any ontological treatment of urban environments must account for both their statics and dynamics (Galton, 2001, 2003; Grenon & Smith, 2004; Worboys, 2001). The ontology is divided into four components

- *continuant* entities, often called *objects*: entities that exist in their completeness at any moment in time, have no temporal parts, but have qualities that might change through time. Examples include roads, vehicles, people, mobile agents, houses, and cities.
- *occurrent* entities, often called *events*: entities that happen, are situated in spacetime and have temporal parts. Examples include journeys, the construction of a house, or dynamic points of interest, such as traffic accidents.
- *processual* entities: entities having some of the characteristics of occurrents, but not anchored in spacetime. Examples include walking (oppose this to a specific walk event). Flows in networks, such as vehicle densities along a transportation link, are often placed in this category (Galton & Worboys, 2005).
- *situational* entities: sometimes called *sites*. These are the spatiotemporal references of entities in the above categories, if they have them. A house is situated in a region of spacetime, a journey might be thought of as a trajectory in spacetime. Depending upon the entity, the temporal component can be more or less important.

When it comes to ontological analysis or conceptual modeling for urban DeSC, not only must the above categories be investigated in the specific scenario presented, but also relationships between them. For example, in an LBS system, a mobile service user (continuant entity) may participate in a commuter journey (occurrent entity) through urban space, which is situated along a specific route in a transportation network (situational entity). It is this detailed and rigorous analysis that can provide the foundation of the computational model of the system.

In the context of DeSC for urban environments, such ontological analysis is paramount as it offers a foundation for querying dynamic systems (Worboys, 2005) and for abstracting movement patterns (Stewart Hornsby & Cole, 2007). Galton and Worboys (2005) identified specifically the application domain of traffic (specifically Ordnance Survey's OS MasterMap® Integrated Transport Network™ layer, ITN) for their conceptual model for dynamic spatial networks.

2.2 Decentralized Spatial Computing

Advances in ad-hoc wireless networking and micro-fabrication have enabled a new way of capturing and processing spatiotemporal information. *Wireless sensor networks* (WSN) — networks of untethered, wireless, battery powered miniaturized computers — monitor their environment by sensing, processing, and communicating information in a collaborative way (Zhao & Guibas, 2004). Recent research activity in the area of WSN has focused on the establishment and maintenance of the network (e.g. Braginsky & Estrin, 2002; Cheng & Heinzelman, 2005), including many ingenious techniques using the spatial characteristics of the network for that purpose (e.g. Karp & Kung, 2000; Mauve, Widmer, & Hartenstein, 2001; Yu, Govindan, & Estrin, 2001). WSN are especially well suited to monitor dynamics in geospace (*geosensor networks* (Nittel, Stefanidis et al., 2004)) and vehicular traffic (*vehicular ad-hoc networks* (VANET) (Kosch, Adler, Eichler, Schroth, & Strassberger, 2006)).

In this chapter we use the term *decentralized* system to specifically refer to a distributed system, where no component of the distributed system 'knows' the entire system state (Lynch, 1996). In decentralized systems, individual elements must cooperate to complete some processing task, but both the task and the data remain distributed throughout the network. Hence, *decentralized spatial computing* aims at the development of algorithms that can operate using purely *local* knowledge, but are still able to monitor geographic phenomena with *global* extents (Estrin, Govindan, & Heidemann, 2000; Laube & Duckham, in press).

For several reasons decentralized (in-network) algorithms are increasingly important to AmSI. First, decentralization increases network scalability and robustness, which is paramount to dynamic urban applications involving potentially thousands or millions of phone users or vehicles. Second, unlike global approaches, decentralized algorithms facilitate fast local updating in dynamic networks, as they do not require hard-to-maintain global consistency. Third, in-network data pre-processing reduces the need for network-wide communication, and hence conserves critically limited energy and bandwidth resources in the WSN. In this chapter we explore the potential of DeSC for providing location based services in a VANET scenario, where energy constraints are negligible as vehicles have engines and hence almost unlimited energy resources. Scalability, robustness, and a constantly changing network topology, however, remain critical VANET issues to be addressed through DeSC.

2.3 Safeguarding Privacy

In pre-internet and pre-database times privacy was safeguarded through the fragmented nature of personal information sources (Rule, McAdam, Stearn, & Uglow, 1980). One's bank knew about personal finances, the police about one's crime record, and the grocer about shopping habits; data integration was impractical, inference nearly impossible. The current inexorable integration of previously disjoined data sources into centralized databases creates omniscient bottlenecks open to fraud.

Recent improvements in mobile computing and location-aware technologies allow for the capture and communication of fine-grained spatiotemporal information about mobile individuals. Trajectories contain information in the form of sensitive personal points of interest and movement patterns (Bettini et al., 2005). Hence, trajectory information may act as *quasi identifier*, allowing the reidentification of anonymized data, just as with a non-spatial social security number. The increasingly dense cyber-infrastructure of our built urban environments increases privacy concerns for unsuspecting users of AmSI.

Conventional approaches protecting the privacy of mobile individuals involve *regulation*, *privacy policy*, and various forms of *data hiding* (see Verykios, Damiani, and Gkoulalas-Divanis (2008) for a comprehensive overview). Regulatory approaches to privacy develop rules to govern fair use of personal information, most importantly legislation (Langheinrich, 2001). Privacy policies rely on trust and stipulate allowed uses of location information (Kaasinen, 2003). Most recently, considerable effort has been put into various strategies of hiding sensitive information, especially sensitive spatial information. Anonymity concerns the dissociation of information about an individual, such as location, from that individual's actual identity. *k*-anonymity — one of the most important anonymity approaches — protects an individual's privacy in that each sensitive release is hidden in at least *k* equally matching individuals (Bettini et al., 2005). Kido, Yanagisawa, and Satoh (2005) hide true users' trajectories by mixing them with synthetic 'fake' ones, so-called 'dummies.' Finally, obfuscation involves the deliberate degradation of the quality of location information (Duckham & Kulik, 2005).

3 Protecting Privacy with DeSC

One of the weaknesses in the previous approaches to location privacy protection reviewed above is that they typically adopt a static model of privacy, aiming primarily to protect an individual's privacy for a specific instant, query, or site. Given the importance of dynamic information in urban environments ('the city in flux'), this research adopts an explicitly spatiotemporal approach to privacy protection. We argue that the disclosure of the odd (static) locational fix is often acceptable, especially as this information rapidly becomes out-of-date. Instead, gathering of trajectory information over time is the most imperative threat to an individual's geospatial privacy, enabling hostile agents to infer spatiotemporal patterns, and analyze or predict behaviors. Thus the focus of this work is less to protect the individual's location at a specific point in time, but rather to protect the trajectory in its entirety as a spatiotemporal entity.

The approach to protecting trajectory information explored in this chapter achieves *spatiotemporal privacy protection* by using DeSC. As we have seen, ubiquitous and location-aware computing environments bring with them privacy threats as they can potential be used to collect automatically more and more detailed spatiotemporal information about an individual's location. However, we argue that DeSC allows the inversion of that process. Where centralized spatial computing architectures make the collation of spatiotemporal data easier, in decentralized architectures sensitive knowledge is 'smeared' across spacetime. Potentially, decentralization ensures no single system component can accumulate detailed knowledge about any individual, and so privacy is protected.

3.1 An LBS scenario

Consider the following scenario for safeguarding privacy in an urban LBS application using decentralization. Multiple mobile individuals are moving through an urban street network, each carrying a mobile location-aware device (like a PDA or wireless sensor node). From this point on we simply refer to the combination of a mobile individual and his or her mobile location-aware device as a *mobile agent*. In a DeSC environment, nearby mobile agents are assumed to be able to communicate with each other using short-range communication (e.g., WiFi or Bluetooth). *Points of interest* (POIs) are also distributed throughout the urban network. These POIs may be static, such as retail stores or coffee shops, or dynamic, like wireless hotspots, meetings, or traffic jams. Whenever an agent encounters a point of interest (POI), the agent's device stores that POI's ID and location. In many cases, the process of identifying POIs may be completely automatic (e.g., a device might automatically identify a wireless hotspot POI via WiFi radio frequency signals, or retail store via RFID). However, in some cases semi-automatic or manual generation of this information is conceivable (e.g., user tagging of points of interest).

Given this scenario, one important task is to be able to answer *k*-nearest neighbor queries, like 'Where is my nearest POI?' Answering such questions is a basic function of conventional centralized LBSs. However, in this work we look at the extent to which DeSC can be used to provide the same function, at the same time as protecting a user's location privacy. An important simplifying assumption in the following discussion and subsequent experiments is that POIs are 'semi-dynamic' in the sense that POI locations are initially unknown and must be discovered by mobile agents, but POIs change with relatively low frequency when compared with the movement of agents. In other words, in the context of our scenario, once a POI has been discovered by an agent it remains valid for a relatively long period of time (when compared with the frequency with which agents move into and out of the system). Such a scenario is suitable for many (but not all) dynamic phenomena (e.g., ad hoc WiFi hotspots, which while dynamic can be usually relied upon to still be active hours or even days after first observed by an agent).

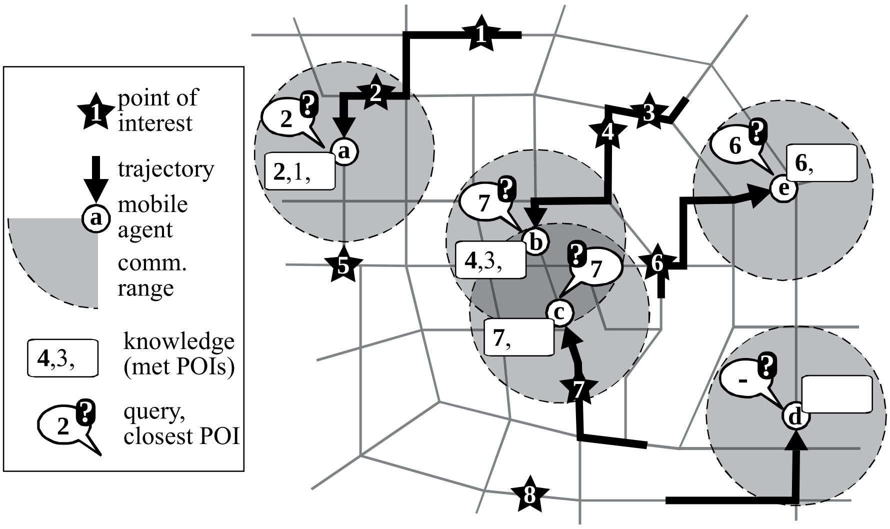

Fig. 1. LBS scenario: mobile agents in an urban road network store information about POIs they have encountered, and can pose nearest neighbor queries to neighboring agents within (*n*-hop) communication range.

Figure 1 shows a simplified example of the basic scenario for five mobile agents *a* - *e*. Mobile agents store information about POIs they have directly encountered (e.g., agent *b* stores location and identity information about POIs 4 and 3). When an agent requires information about the nearest POI it can query its own stored data as well as query any neighbors within range (e.g., agent *b* can also query agent *c* for its closest POI). For clarity, the scenario in Figure 1 shows a disconnected agent network, where only 1-hop communication at most is possible. However, in more connected networks, the nearest neighbor query may involve multiple hops, potentially querying the entire network of agents, maximum hops and network connectivity allowing.

Thus, over time in our scenario agents moving around the urban environment will discover more and more of the environment, enabling the system as a whole to answer nearest-neighbor queries more and more accurately; but at the same time agents will reveal information about where they are located when they pose a query, potentially reducing their location privacy.

3.2 Experimental methodology

The performance of a decentralized, location-based nearest-POI service was explored empirically using a series of simulations. The goal of these simulations was to investigate the extent to which such decentralized LBS can provide useful services at the same time as effective spatiotemporal privacy protection. The simulation was programmed using a combination of Java and Oracle spatial DBMS, linked using JDBC (Java data-base connectivity) APIs. The experiments used a small part of Ordnance Survey's ITN data set for downtown Southampton, on the southern coast of the UK. Moving toward data models and databases that are capable of supporting AmSI is an important research problem for national mapping agencies like Ordnance Survey, especially within the context of next-generation intelligent transportation systems (ITS).

Each experiment was initialized with 100 randomly positioned POIs. A total of 100 agents was also randomly positioned in the network, initially with no knowledge of the location of the POIs. At each time step, agents move to adjacent nodes in the network. Any POIs encountered by (i.e., co-located with) an agent results in that POI's ID and location being stored in that agent's local memory. In this simplified simulation, no agents or POIs are ever destroyed or leave the simulation, and no agents or POIs are created after initialization of the simulation.

Over time, the mobile agents explore more of the network, discovering more POIs. Figure 2 shows a typical example of the state of the system after 50 time steps, with darker shaded circles indicating agents with knowledge of the location of more POIs. Agents can choose to engage in peer-to-peer communication with other nearby agents. The simulation allows a range of communication radii to be used, and single or multi-hop communication. In this way, agents can choose to share information about POIs they have encountered, or query nearby agents about the nearest POIs.

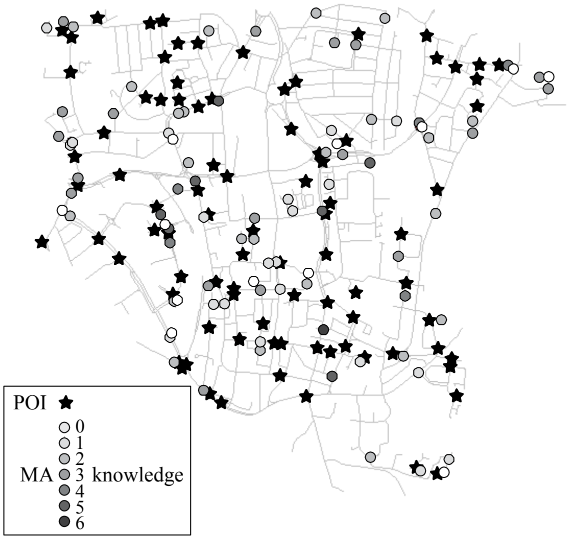

Fig. 2. Example snapshot after *t* = 50 time steps of the state of a simulation, with mobile agents (MA, dots) moving around the road network of central Southampton, UK. Darker shading of agents indicates an agent has knowledge of the location of more POIs (stars). Data is OS MasterMap® Integrated Transport Network™ Layer Ordnance Survey © Crown Copyright. All rights reserved.

The following sections describe four increasingly sophisticated experiments that begin to uncover the behavior of a decentralized system for safeguarding trajectory privacy across a range of experimental parameters.

4 Experiment #1: Quality of service vs. level of privacy

The first experiments evaluated the levels of privacy and quality of service attainable by mobile agents using a decentralized nearest POI query. Initial simulations explored the situation where a *query agent* queried all other agents within its communication neighborhood at each time step for the query agent's nearest POI. The query agent and its neighboring agents used their own knowledge of POIs, if any, to compute the answer to the nearest neighbor query and return the POI location and network distance to the query agent. The best answer (in terms of shortest distance to POI) was then selected from all the responses. It was assumed that all agents possessed complete knowledge of the static road network (but, note, not the semi-dynamic POI locations). Although this is a somewhat simplifying assumption, it is not unrealistic as today many mobile devices currently store large amounts of (static) transportation network data.

In this dynamic scenario it is not possible to guarantee *exact* answers to queries. For example, from Figure 1, agent *b* may be closest to POI 6. Because none of the agents within communication range of *b* at the time of query have encountered POI 6, agent *b* may receive a sub-optimal nearest neighbor query response of POI 7. Given that the POIs are assumed to be (semi-)dynamic, it would similarly not be possible to guarantee exact answers to queries using a centralized approach (for example, where discovered POIs are reported back to a central server; in such a case POI 8 would still remain undiscovered). However, unlike a centralized approach, using decentralized queries enables a balance between quality of service and level of privacy to be struck. If agents choose to reveal their location information to more neighbors (e.g., using multi-hop communication or larger communication distances), then the likelihood of receiving an optimal nearest neighbor response increases, but leads to lower levels of privacy. Conversely, if an agent decides to reveal its location information to fewer neighbors, then the levels of privacy for that agent is expected to increase but at the cost of lower quality of service (i.e., fewer optimal query responses). Striking an acceptable balance between level of privacy and quality of service is one of the key goals of a good privacy protection system.

Figure 3 shows a typical example of how levels of privacy and quality of service (averaged across 100 simulations) vary for agents over 100 simulation time steps, referred to as a *QoS/LoP signature*. In Figure 3, the quality of service is measured in terms of the inaccuracy in query responses. Two different measures of inaccuracy are used: the fraction of agents who do not receive the optimal query response, and the percentage of agents who do not receive any response at all. As would be expected, at the beginning of the simulations, levels of inaccuracy are extremely high, since the agents have not yet explored much of the environment, and so know relatively little about the POIs. However, as the simulation proceeds, quality of service rapidly improves, indicated by rapidly dropping levels of inaccuracy. At the end of the simulation, approximately 90% of agents are receiving the optimal query response (i.e., the closest POI).

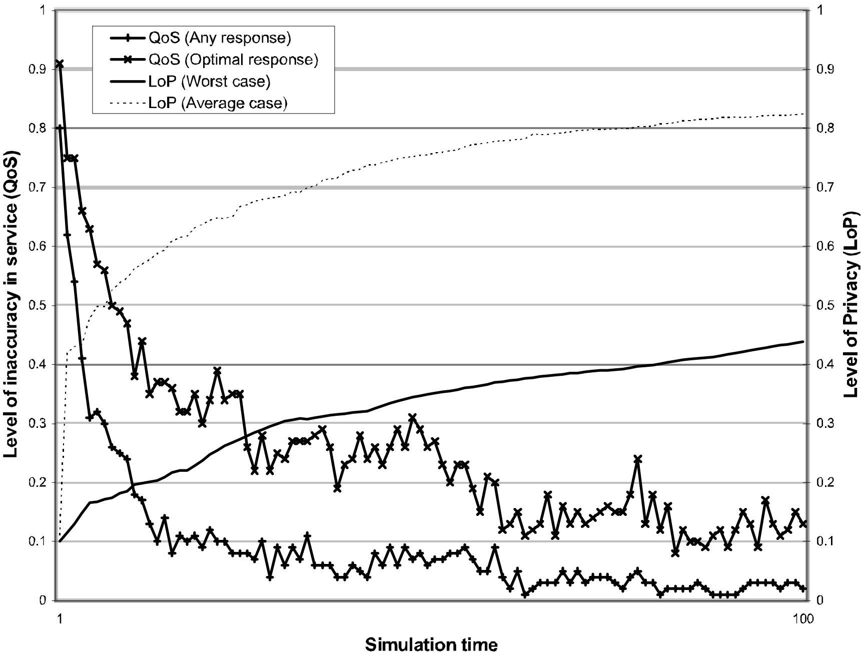

Fig. 3. Typical QoS/LoP 'signature', showing increasing average and worse case levels of privacy over simulation time, and increasing quality of service, in terms of decreasing inaccuracy in query responses.

The level of privacy of agents is also quantified with two different measures. The average level of privacy indicates what percentage of an agent's trajectory on average is unknown to any other arbitrarily chosen single agent. Figure 4 illustrates the general idea behind using knowledge of trajectories as a basis for measuring the level of privacy. In the figure, assuming agent *b* communicates its location to agent *c* during the entire duration they are in direct 1-hop communication range, about 70% of agent *b*s trajectory is unknown to *c* (a level of privacy for *b* of 0.7 with respect to *c*). The average level of privacy is initially very low, since after only a few time steps knowledge of only one or two locations for a query agent is likely to constitute a high percentage of that agent's entire trajectory, and so low levels of privacy. However, as simulation time allows the system to equilibrate, the level of privacy increases asymptotically. At the end of 100 time steps, an arbitrarily chosen node will, on average, only know less that 20% of a query agent's trajectory (indicated by an average level of privacy of more than 80%).

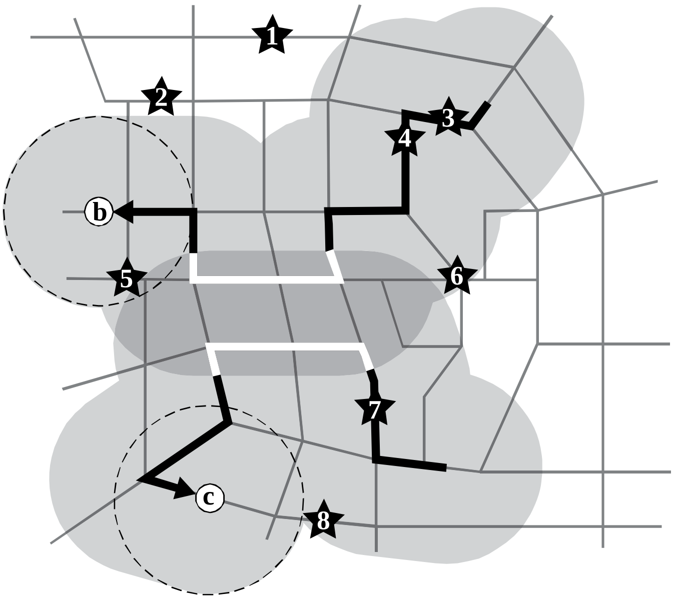

Fig. 4. Level of privacy is measured as percentage of an agent's trajectory that is not known to other agents. Agent *b* does not know about ~60% of *c*s trajectory (*c* has privacy level of 0.6 with respect to *b*; by contrast, *c* does not know about ~70% of *b*s trajectory (*b* has privacy level of 0.7 with respect to *c*).

While average level of privacy does provide a good picture of the degree to which trajectory information about a mobile agent is 'smeared' across the network, it can be argued that it is a poor overall measure of location privacy, since only *one* node is needed to potentially result in a breach in privacy. To better reflect the potential privacy risk of one agent invading another's privacy, the second measure of level of privacy is the worst case privacy which indicates what percentage of a query agent's trajectory on average is unknown to the agent that knows *most* about that query agent. Although the worst case level of privacy is necessarily lower than the average case, the pattern is still the same, showing steadily increasing levels of privacy.

Discussion. Figure 3 shows that it is possible to achieve a balance of quality of service and level of privacy using decentralized spatial computing techniques. The levels of privacy provide a reflection of how information about an agent's location is 'smeared' across spacetime. While 100% accuracy in quality of service cannot be guaranteed, relatively high qualities of service can be achieved (and even for a centralized LBS, quality of service would never be perfect either, since typically some POIs may by chance remain undiscovered by any agent).

Most importantly, when explicitly protecting trajectory privacy and not location privacy, the suggested decentralized peer-to-peer query procedure allows individual agents the spatiotemporal accumulation of knowledge for a better QoS without compromising their privacy. In other words, when protecting trajectory privacy, levels of privacy and quality of service *both* can improve over time. This contrasts to most approaches protecting location privacy, where the collation of knowledge over time in centralized architectures leads to worsening privacy.

5 Experiment #2: Communication network effects

The results presented in section 4 are typical of simulation results across a range of simulation parameters. Two such important parameters are related to the communication network used for locally querying nearby agents in a decentralized way. The communication range of agents and number of 'hops' used to propagate queries through the network will affect both the balance of quality of service and level of privacy. Larger communication radii and more hops will tend to lead to improved quality of service (since it extends the communication neighborhood of a node, resulting in more nodes contributing their knowledge of the environment to the query response) as well as decreased levels of privacy (since more nodes will be informed of the location of a query agent).

The QoS/LoP signature in Figure 3 showed the results for a simulation using a communication radius of about 1/16 of the diameter of the study area (1/16 being ~200m), and using 1-hop communication for queries. In other words, in Figure 3 only agents that are at most 200m from a query agent will respond to a query. Figure 5 summarizes the effects of changing communication radius on quality of service. A range of communication radii were tested, measured as a proportion of the size of the overall study area (so for example, a communication radius of 1/2 translates into approximately 1600m, half the diameter of the study area; 200m translates into a communication radius of approximately 1/16).

As expected, quality of service increases (i.e., service inaccuracy decreases) asymptotically across all communication radii over simulation time, with larger communication radii generally leading to higher service qualities. It is noticeable that increasing the communication radius to above about 1/4 of the study area diameter leads to no further perceptible improvements in quality of service. In other words, at communication radii above 1/4 of the study area size, no additional information relevant to nearest POIs queries is being found.

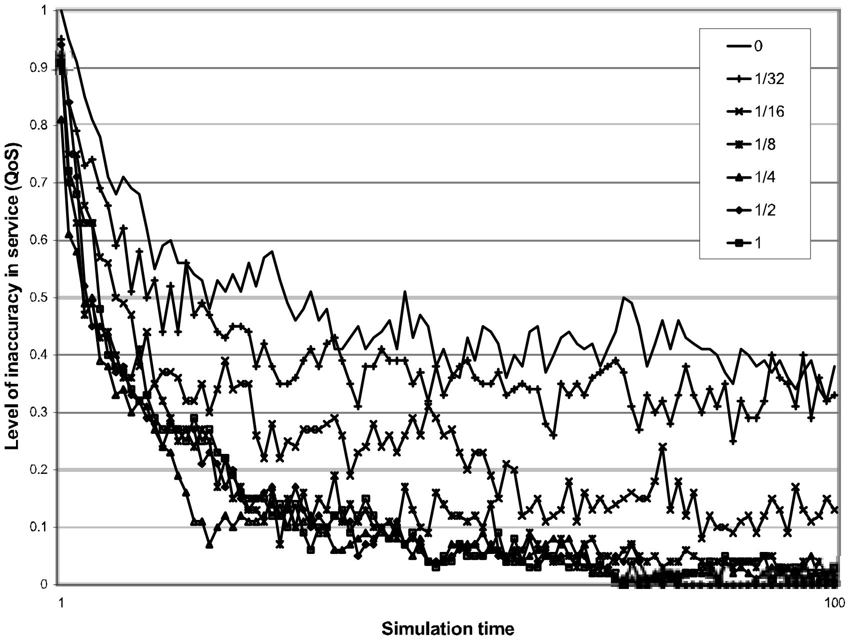

Fig. 5. Effects of communication radius on quality of service. Communication radius is measured as a proportion of the size total study area, i.e., communication radius of 1 means agents can communicate directly with all other agents in the study area, radius of 0 means agents do not communicate.

Discussion. The effects of increasing the number of hops for queries is similar to increasing the communication range (since on average doubling the communication range means approximately as many nodes will be queried as by doubling the hop limit). Consequently, it is not surprising that similar results to those in Figure 5 were obtained from similar studies of increasing hops in queries. The key message of Figure 5 is that most of the information relevant to a nearest POI query can be expected to reside in close spatial proximity to the query agent. Thus increasing hops or communication radius leads to diminishing returns in terms of increased quality of service. Conversely, level of privacy rapidly decreases for increasing communication radius and hops, such that for any communication radius of above 1/4 of the study area diameter or hop count above 3, there almost always exists at least one agent that has near-complete knowledge of a query agent's trajectory (i.e., the query agent has zero privacy). Thus, in decentralized LBS, the communication network is critical to maintaining an acceptable balance of level of privacy and quality of service. Communication needs to be restricted enough to prevent privacy being compromised, at the same time as enabled enough to ensure reasonable quality of service.

Finally, it is also noticeable that in Figure 5 a communication radius of 0 (i.e., no communication, agents can only query their own database of POIs they have already seen) still leads to relatively high qualities of service, of around 60% optimal answers. This can be ascribed to the random walk movement behavior of the agents, which means they tend to stay and thoroughly explore relatively small, spatially constrained regions of the environment. The following experiment addresses this shortcoming.

6 Experiment #3: Goal directed movement

As discussed in the previous section, the movement regime of agents can affect the observed quality of service and level of privacy. Random walk, used in the previous experiments, is not a realistic movement regime for most mobile humans accessing location-based services. To address this issue, the experiments were repeated using goal-directed movement, where agents move from their current location to a randomly selected destination using the shortest path. When an agent reaches its destination in our simulation, it is immediately re-tasked with a new randomly selected destination, to which it again moves along the shortest path. The effect of changing the movement behavior to goal-directed is expected to both increase level of privacy (since a query agent is expected to meet and reveal its location to a wider range of other agents less frequently) and decrease quality of service (since in general agents in a locality are less likely to have thoroughly explored and identified all the POIs in that area).

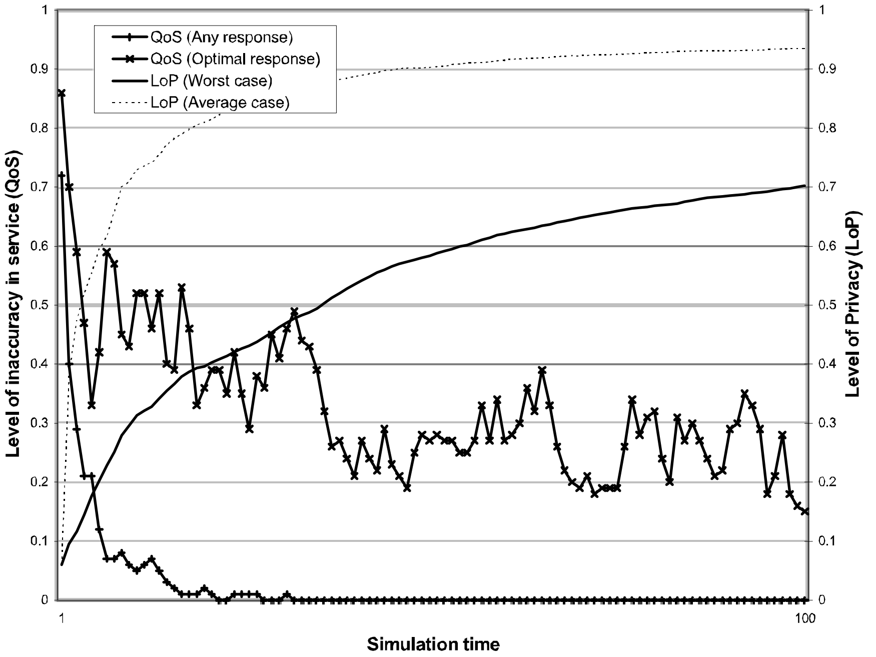

Fig. 6. QoS/LoP signature for same scenario as Figure 3, except where agents perform goal-directed movement rather than random walk.

Figure 6 shows the signature from the same simulation scenario as Figure 3 (communication radius 1/16 of the study area diameter, and 1-hop communication), except where agents are using goal-directed movement instead of random walks. As expected, the levels of privacy are indeed increased, rising to 0.7 in the worst case (i.e., at the end of the simulation, there exists no agent that knows more than 30% of the trajectory of a query agent), and the quality of service is decreased somewhat when compared with agents performing random walks. It is interesting to note, however, that contrary to expectations the quality of service in terms of query agents who receive *some* answer (albeit not necessarily the optimal answer) actually *increases*. This is presumably because goal-directed movement results in greater mixing of agents, also ensuring good mixing of knowledge. Conversely, random walk tends to allow clustering of agents, and occasionally of ignorant agents who happen to have no knowledge of nearby POIs.

Discussion. Since goal-directed movement tends to increase mixing and aid spacetime ‘smearing’ of an agent’s location information, agents engaged in such movement regimes can afford to reveal more about their location while still maintaining the same overall level of privacy. Experiments with a range of network communication parameters revealed that using 1-hop communication with a communication radius of 1/8 of the study area yielded some very favorable results, with worst case levels of privacy of 0.5; average case privacy of 0.15; more than 80% of agents receiving an optimal answer; all agents receiving some answer; and more than 90% of agents receiving a nearest POI that was less than 50% further away than the optimal answer (i.e., relatively good suboptimal-answers). While the highly simplified simulation setup means that these values do not have any real meaning outside the simulation, this result does provide encouragement that by varying the decentralized query parameters it is possible to achieve high levels of privacy in concert with high qualities of service.

7 Experiment #4: Push and pull queries

In all the simulations discussed thus far, agents query using a 'pull' strategy, where information is only ever exchanged between agents when a query is generated. However, it is also possible to design decentralized privacy protection systems that utilize 'push' strategies, where information is opportunistically pushed to communication neighbors in case required at a later stage.

In contrast to the discreet pull strategy in Figure 1, push is a more loquacious. Using the push strategy, mobile agents synchronize their knowledge of POIs with any nearby agents they 'meet' (i.e., agents that move into each other's *n*-hop neighborhood). Figure 7 illustrates the push strategy for the same mobile agents, POIs, and trajectories as Figure 1. The key difference between the figures is that agents in Figure 7 'push' information about POIs whether or not an explicit query has been received. As a result, mobile agents can accumulate information about remote POIs they have not directly encountered (e.g., agent *c* 'hears' about POIs 3, 4, and 6 from agent *b*, who in turn hears about POI 6 from agent *e*).

In a pure push strategy, agents who later require a response to a nearest neighbor query can simply query their own locally stored data about POIs without querying nearby nodes. The push strategy is similar to the concept of information dissemination in mobile ad-hoc geosensor networks (and specifically the 'flooding' strategy) explored in more detail in Nittel, Duckham, & Kulik (2004).

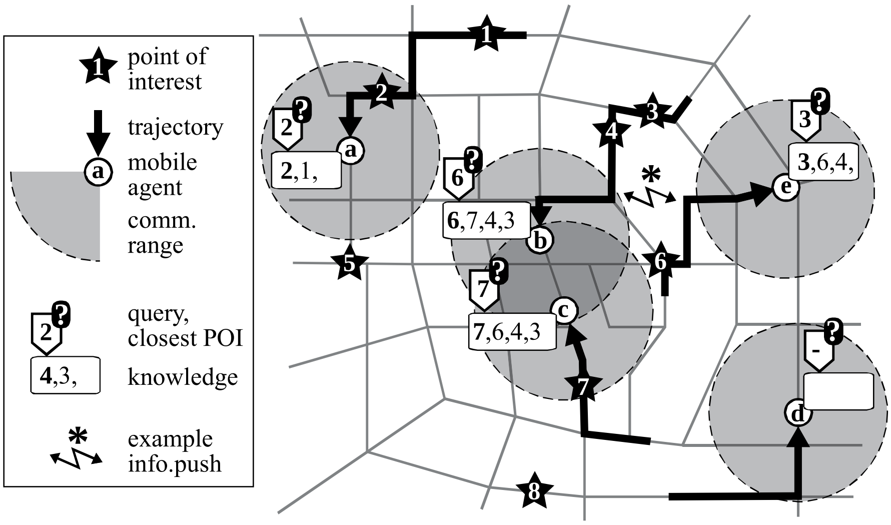

Fig. 7. Push strategy: mobile agents in an urban road network store information about POIs they have encountered and share this information with other neighboring agents within (*n*-hop) communication range. Nearest neighbor queries can then be submitted to the agent's local POI database (cf. Figure 1).

Figure 8 illustrates the typical effects of adopting a push strategy, like that explored in Nittel et al. (2004), upon the QoS/LoP signature. Figure 8 uses the same basic simulation parameters as Figure 6 (i.e., goal-directed movement, communication range of ~200m). Using the push strategy, agents exchange at every opportunity all their knowledge about the locations of POIs with any other agent within direct 1-hop communication range. Unlike the pull strategy, when a query agent subsequently requires information about the nearest POI, it merely queries it's own stored POI data.

Figure 8 illustrates that the push strategy performs better in terms of QoS that the corresponding pull strategy. This is to be expected, as the push strategy leads to much more widespread dissemination of POI information, ensuring that agents are much more likely to receive high quality responses to queries. Indeed, Figure 9 shows how using a push strategy can result in improved QoS almost irrespective of communication radius (since using a push strategy, all agents will 'hear' about newly discovered POIs after a relatively small number of hops).

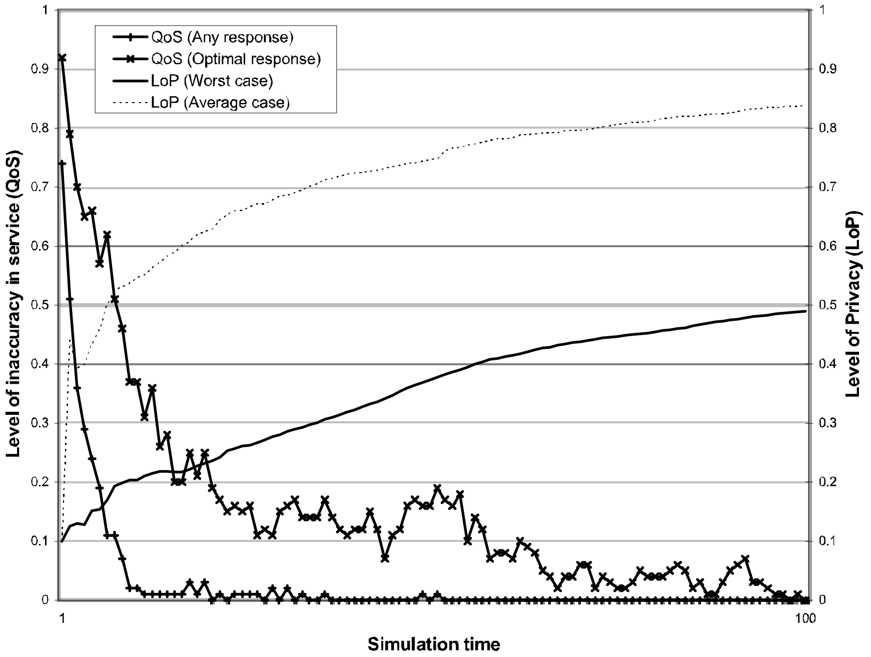

Fig. 8. QoS/LoP signature for same scenario as Figure 6, except where agents use push rather than pull strategy.

However, also as might be expected the push strategy leads to lower levels of privacy than the pull strategy, as it requires that agents reveal information about their location much more frequently (i.e., whenever they meet another agent). In practice, the discrepancy between the levels of privacy achievable with a push and pull strategies is likely to be much greater than suggested by Figures 8 and 6. Agents using a pull strategy are unlikely to require near-continuous responses to queries as in our simulations, needing instead to make occasional pull queries. Thus in practice, the levels of privacy achieved by pull queries are expected to be much higher than in our simulations. However, because it is not known in advance what information will be subsequently required, there are no such obvious privacy optimizations that can be adopted using a push strategy.

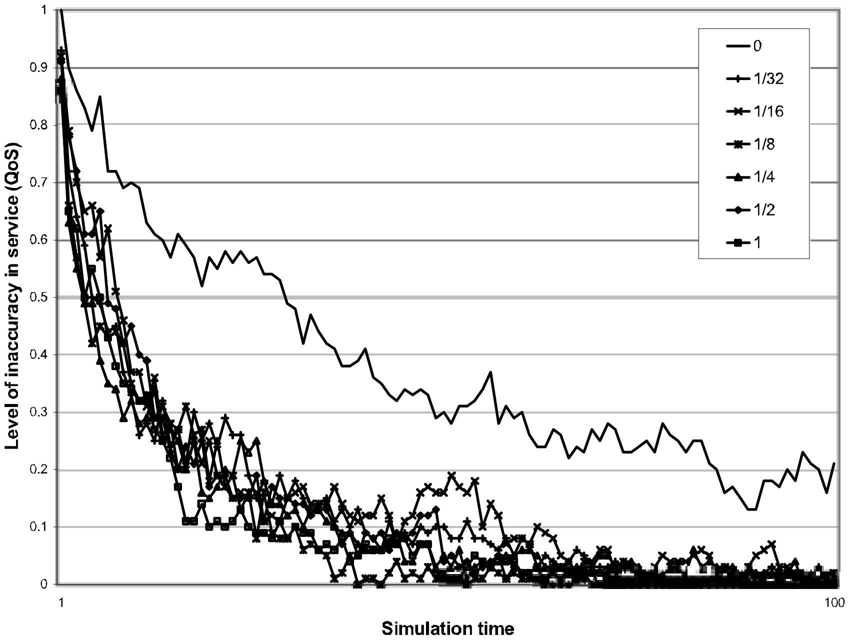

Fig. 9. Effects of communication radius on quality of service for push strategy (cf. Figure 5).

Discussion. The results clearly indicate that while push strategies can improve QoS, equivalent pull strategies are expected to provide better privacy protection. While the focus of this work is on privacy protection, an important practical issue that is not addressed by these simulations is the communication and computational overhead of the different strategies. Push strategies generally lead to much higher volumes of data being stored and exchanged, placing strong demands on resources such as communication bandwidth, data storage space, and battery power. While these issues are not as important in most mobile computing environments as they are in, for example, wireless sensor networks, they are still a potential threat to system scalability.

8 Conclusions and Outlook

This chapter promotes the vision of ambient spatial intelligence (AmSI) for urban environments. Decentralized spatial computing (DeSC) is presented as a key computational strategy enabling AmSI. It is argued that decentralization not only copes with the highly dynamic computing systems underlying AmSI, but even offers additional benefits not easily accessible with comparable centralized solutions. The additional benefit from decentralization addressed in this chapter is safeguarding privacy of LBS users. For a decentralized LBS the balance of quality of service and level of privacy has been investigated with a comprehensive set of consecutive experiments covering various network communication parameters (one-hop vs. multi-hop communication, variable communication ranges) and motion regime parameters (random vs. goal directed walk). The experiments revealed that for carefully chosen network communication parameters the decentralized approach indeed allowed protection of privacy whilst maintaining reasonable quality of service.

One identified advantage of a decentralized LBS is its inherent affinity for dynamics. The same motivating arguments as presented for 'semi-dynamic' POIs, certainly hold for even more dynamic scenarios, when POI change with high turnover rates. When quick response to a constantly changing topology is prime, then the intimate relation of proximity and decentralization becomes most obvious. Accepting that nearby changes are probably most relevant to other service users nearby, what could be more efficient and effective than exchanging locally relevant information only with other agents nearby? The chosen example of safeguarding privacy through service decentralization illustrates that accepting the challenge of in-network data processing not only presents additional complexity but potentially also offers benefits. Not only does local information exchange limit the depletion of global network resources, but, as has been shown in the experiments presented in this chapter, it may also safeguard privacy as agents receive the service they want with only local disclosure of their potentially sensitive location information. Clearly, it is exactly this application layer in the otherwise technology driven area of AmSI where GIScience has to make its contribution, as has been shown with exploiting the spatiotemporal nature of movement for safeguarding privacy.

DeSC and geosensor networks have been widely explored for environmental monitoring, largely focusing on enabling spatial applications under the harsh technological constraints of WSN. Well defined monitoring tasks — such as for example tracking an evolving contamination area — proved to be very useful for inaugural research on DeSC. Such initial work on DeSC addressed well known GIScience challenges, including the complexity of spatial data, uncertainty, and interoperability. Additionally, the new technologies allowing DeSC also revealed the need for in-network data processing and decentralization which in turn presents a set of new exciting challenges. First, highly dynamic p2p computing steps alongside client-server architectures. Second, decentralization promotes the notion of local knowledge as opposed to global knowledge. Consequently, decentralization and partial data processing inherently involves incomplete knowledge, requiring in turn the use of heuristics and approximation solutions. In general, these old and new challenges to DeSC also apply for urban applications. There are, however, some more challenges emerging specifically in urban DeSC.

The number of agents in AmSI is potentially much larger than a few thousand nodes deployed in an ecological monitoring network. Just imagine AmSI applications involving RFID tags on retail products. Quite in contrast to environmental applications, the networks used for urban AmSI will rarely be deployed explicitly for a DeSC application and rather make use of existing cyber-infrastructure (e.g. buddy tracking in a cell phone network). Hence, the secondary DeSC application will have much less influence on system configuration and tasking. Hence, urban DeSC applications will rather aim at exploiting given constraints than relying on lofty assumptions. Finally, urban WSNs are expected to be highly heterogeneous, potentially including nodes as different as radio relays, vehicle board-computers, handsets and RFID tags all together. DeSC for urban applications will hence face the difficulty of processing even larger and more heterogeneous data than environmental geosensor applications.

To conclude, we identify four main research and development topics for DeSC in the urban context:

- With respect to decentralized privacy protection in LBS, a further investigation of hybrid push-pull approaches, potentially reaching an optimized QoS and LoP.
- Further exploration of DeSC for safeguarding privacy in mobile communication applications (e.g., buddy tracking, child watch).
- Relaxation of the simplifying assumption of homogeneous single-purpose networks and investigation of DeSC applications for heterogeneous networks.
- The exploration of decentralized spatial data mining techniques for mobile WSN, especially suited for very large and very heterogeneous systems emerging urban AmSI applications.

Acknowledgments Patrick Laube is funded by the Australian Research Council (ARC), Discovery grant DP0662906 and the ARC Research Network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP). Matt Duckham's research is supported by the Australian Research Council under ARC Discovery Grant DP0662906. The authors would furthermore like to thank Ordnance Survey of Great Britain.

References

Bettini, C., Wang, X., & Jajodia, S. (2005). Protecting Privacy Against Location-Based Personal Identification. In W. Jonker & M. Petkovic (Eds.), *Secure Data Management* (Vol. 3674, pp. 185-199). Heidelberg: Springer.

Braginsky, D., & Estrin, D. (2002). Rumor routing algorthim for sensor networks. In *Proceedings of the 1st ACM international workshop on wireless sensor networks and applications* (pp. 22-31). Atlanta, GA: ACM Press.

Cheng, Z., & Heinzelman, W. B. (2005). Flooding strategy for target discovery in wireless networks. *Wireless Networks, 11*, 607-618.

Dillenburg, J. F., Wolfson, O., & Nelson, P. C. (2002). *The Intelligent Travel Assistant.* Paper presented at the The IEEE 5th International Conference on Intelligent Transportation Systems.

Dobson, J. E., & Fisher, P. F. (2003). Geoslavery. *IEEE Technology and Society Magazine, 22*(1), 47-52.

Duckham, M., & Kulik, L. (2005). Simulation of obfuscation and negotiation for location privacy. In *Spatial Information Theory ({COSIT} 2005)* (Vol. 3693, pp. 31-48). Heidelberg: Springer.

Duckham, M., Nittel, S., & Worboys, M. F. (2005). *Monitoring dynamic spatial fields using responsive geosensor networks.* Paper presented at the ACM GIS.

Estrin, D., Govindan, R., & Heidemann, J. (2000). Embedding the Internet - Introduction. *Communications of the ACM, 43*(5), 38-41.

Galton, A. (2001). Space, time, and the representation of geographical reality. *Topoi-An International Review of Philosophy, 20*(2), 173-187.

Galton, A. (2003). Desiderata for a spatio-temporal geo-ontology. In *Spatial Information Theory: Foundations of Geographic Information Science (COSIT 2003)* (Vol. 2825, pp. 1-12). Heidelberg: Springer.

Galton, A., & Worboys, M. (2005). Processes and events in dynamic geo-networks. In M. A. Rodrìguez, I. F. Cruz, M. J. Egenhofer & S. Levashkin (Eds.), *Proceedings of First International Conference on Geospatial Semantics* (Vol. 3799, pp. 45-59). Heidelberg: Springer.

Greenfield, A. (2006). *Everyware: the dawning age of ubiquitous computing*. Berkeley: New Riders Press.

Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation, 4*(1), 69-103.

Kaasinen, E. (2003). User needs for location-aware mobile services. *Personal and Ubiquitous Computing, 7*(1), 70-79.

Karp, B., & Kung, H. T. (2000). *GPSR: Greedy perimeter stateless routing for wireless networks.* Paper presented at the Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, Boston, MA.

Kido, H., Yanagisawa, Y., & Satoh, T. (2005). *An anonymous communication technique using dummies for location-based services.* Paper presented at the International Conference on Pervasive Services (ICPS '05).

Kosch, T., Adler, C. J., Eichler, S., Schroth, C., & Strassberger, M. (2006). The scalability problem of vehicular ad hoc networks and how to solve it. *Wireless Communications, IEEE [see also IEEE Personal Communications], 13*(5), 22-28.

Langheinrich, M. (2001). Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. In *Ubicomp 2001: Ubiquitous Computing* (pp. 273-291).

Laube, P., & Duckham, M. (in press). Decentralized spatial data mining for geosensor networks. In H. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (2nd ed.). London: Taylor & Francis.

Lynch, N. (1996). *Distributed Algorithms*. San Mateo, CA: Morgan Kaufmann.

Mauve, M., Widmer, A., & Hartenstein, H. (2001). A survey on position-based routing in mobile ad hoc networks. *Network, IEEE, 15*(6), 30-39.

Nittel, S., Duckham, M., & Kulik, L. (2004). *Information dissemination in mobile ad-hoc geosensor networks.* Paper presented at the Geographic Information Science, GIScience, Heidelberg.

Nittel, S., Stefanidis, A., Cruz, I., Egenhofer, M. J., Goldin, D., Howard, A., et al. (2004). Report from the First Workshop on Geo Sensor Networks. *ACM SIGMOD Record, 33*(1).

Rule, J., McAdam, D., Stearn, L., & Uglow, D. (1980). *Politics of Privacy*.

Stewart Hornsby, K., & Cole, S. (2007). Modeling Moving Geospatial Objects from an Event-based Perspective. *Transactions in GIS, 11*(4), 555-573.

Verykios, V. S., Damiani, M. L., & Gkoulalas-Divanis, A. (2008). Privacy and Security in Spatiotemporal Data and Trajectories. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, Data Mining and Privacy* (pp. 213-240). Heidelberg: Springer.

Winter, S., & Nittel, S. (2006). Ad hoc shared-ride trip planning by mobile geosensor networks. *International Journal of Geographical Information Science, 20*(8), 899 - 916.

Worboys, M. F. (2001). Modelling Changes and Events in Dynamic Spatial Systems with Reference to Socio-Economic Units. In A. U. Frank, J. Raper & J. P. Cheylan (Eds.), *Life and motion of socio-economic units* (Vol. 8, pp. 129-137). London: Taylor \& Francis.

Worboys, M. F. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science, 19*(1), 1-28.

Worboys, M. F., & Duckham, M. (2006). Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science, 20*(10), 1087-1108.

Worboys, M. F., & Hornsby, K. (2004). From objects to events: GEM, the geospatial event model. In M. J. Egenhofer, C. Freksa & H. Miller (Eds.), *3rd International Conference on Geographic Information Science (GIScience 2004)* (Vol. 3234, pp. 327-343). Heidelberg: Springer.

Yu, Y., Govindan, R., & Estrin, D. (2001). *Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks* (No. UCLA/CSD-TR-01-0023): UCLA Computer Science Department.

Zhao, F., & Guibas, L. J. (2004). *Wireless Sensor Networks - An Information Processing Approach*. San Francisco, CA: Morgan Kaufmann Publishers.