Simulation of Obfuscation and Negotiation for Location Privacy

Matt Duckham[1] and Lars Kulik[2]

[1] Department of Geomatics,
University of Melbourne, Victoria, 3010, Australia
mduckham@unimelb.edu.au
[2] Department of Computer Science and Software Engineering,
National ICT Australia Victoria Laboratory,
University of Melbourne, Victoria, 3010, Australia
lkulik@cs.mu.oz.au

Abstract. Current mobile computing systems can automatically sense and communicate detailed data about a person's location. Location privacy is an urgent research issue because concerns about privacy are seen to be inhibiting the growth of mobile computing. This paper investigates a new technique for safeguarding location privacy, called *obfuscation*, which protects a person's location privacy by degrading the quality of information about that person's location. Obfuscation is based on spatial imperfection and offers an orthogonal approach to conventional techniques for safeguarding information about a person's location. Imprecision and inaccuracy are two types of imperfection that may be used to achieve obfuscation. A set of simulations are used to empirically evaluate different obfuscation strategies based on imprecision and inaccuracy. The results show that obfuscation can enable high quality of service in concert with high levels of privacy.

1 Introduction

Pervasive location-aware systems offer a new class of personalized information based services due to their ability to continuously monitor, communicate, and process information about a person's location with a high degree of spatial and temporal precision and accuracy. Those systems are able to collate large amounts of location information into user profiles that provide a complete history of a user's movements. Although user profiles can be used beneficially to offer highly personalized services to a user, location information is sensitive personal information that needs to be protected.

The protection of location privacy is a crucial factor for facilitating the widespread use of location-aware technologies. Privacy issues are considered to be one of the key research challenges in location-aware computing [11]. Unrestricted access to location information is associated with a range of potential negative effects, including *location-based "spam,"* where businesses could exploit the knowledge of a person's location for unsolicited product marketing; decreased *personal safety*, for example from stalking or assault; and *intrusive inferences*, where a person's political views or individual preferences are inferred from their location (see [7, 13, 9]).

A.G. Cohn and D.M. Mark (Eds.): COSIT 2005, LNCS 3693, pp. 31–48, 2005.

1.1 Obfuscation and Automated Negotiation

Our approach to protecting location privacy aims to offer high quality location-based services based on imperfect spatial information. The use of spatial imperfection for privacy is a novel approach suggested in [2], which enables a person to access information relevant to his or her spatial position while safeguarding personal location privacy by revealing the least possible information about that position. We call the process of degrading the quality of information about a person's location, with the aim of protecting that person's location privacy, *obfuscation*.

Individuals using obfuscation should be able to balance their desired level of privacy against their desired quality of location-based service (LBS). In this paper we investigate using automated negotiation in order to achieve a satisfactory balance of the level of privacy and the quality of service. Higher levels of location privacy are likely (although not guaranteed) to entail lower levels of quality for LBS. Achieving the best balance between location privacy and quality of service lies at the heart of successful negotiation strategies. The idea behind automated negotiation is to facilitate practical mechanisms that location-based service providers can implement to attain effective obfuscation based on user preferences, without the need for high levels of explicit user interaction with the obfuscation system.

Obfuscation requires the ability to offer high quality LBSs based on *imperfect* spatial information. This approach is motivated by the initial work on navigation algorithms under spatial imprecision [3], which has developed strategies for providing navigation services to an individual without knowledge of that individual's precise location. Obfuscation allows the identity of a person to be revealed, but that person's location to be hidden. This contrasts with more conventional approaches to location privacy, where a person's identity is hidden but his or her location is revealed (see section 2).

1.2 Imperfect Spatial Information

The use of imperfect spatial information is a key concept in obfuscation. In the literature at least three types of imperfection in spatial information are identified: (1) *imprecision*, which refers to a lack of specificity in information, (2) *inaccuracy*, which is a lack of correspondence between information and reality, and (3) *vagueness*, often characterized by the existence of boundary cases in information [15]. An inaccurate description of an agent's location means that the agent's actual location differs from the conveyed location: the agent is lying about its current location. An imprecise position description could be a region including the actual location (instead of the location itself). A vague description could involve linguistic terms, for example that the agent is "far" from a certain location.

In this article, we compare strategies based on imprecision and inaccuracy to obfuscate an individual's location. We focus on nearest point of interest (POI) queries, which are location-based proximity queries such as "Where is my nearest sushi restaurant?" In particular, we address the question to what extent can a high quality of service be combined with high levels of privacy for nearest POI queries. Important aspects discussed in this paper are the impact of the shape of an obfuscation region and the implications of its initial size for negotiating location privacy.

The paper is structured as follows: Section 2 compares an obfuscation-based approach with current approaches for location privacy. The model for negotiating location privacy is introduced in Section 3. The simulation experiments are explained in Section 4 and their results are given in Section 5. Section 6 concludes the paper and outlines further research.

2 Background

Research into how to safeguard an individual's privacy is becoming an urgent issue in pervasive computing. Most approaches to protecting (location) privacy fall into three categories: *regulation*, *privacy policies*, and *anonymity*. Each of these approaches plays an important role in providing a complete solution to location privacy, but each approach also has its limitations.

Regulatory approaches to privacy develop rules to govern fair use of personal information, including legislation. Langheinrich [10] gives an overview of the history and current status of privacy legislation and examines international fair information practices. However, regulations often lag behind new technology and ideas, and apply "across the board" making them difficult to tailor to specific contexts that may arise.

Privacy policies stipulate allowed uses of location information. Kaasinen [9] surveys policy-driven approaches to location privacy. Privacy policies rely on trust and, therefore, are vulnerable to inadvertent or malicious disclosure of private information.

Anonymity concerns the dissociation of information about an individual, such as location, from that individual's actual identity. A special type of anonymity is *pseudonymity*, where an individual is anonymous, but maintains a persistent identity (a pseudonym) [12]. Although anonymity techniques are fundamental to privacy protection they have limitations, especially in spatial application domains. A person's identity can often be inferred from his or her location, so anonymity and pseudonymity are vulnerable to data mining [4,1]. Further, anonymity presents a barrier to authentication and personalization, which are required for a range of applications [8, 10].

Obfuscation offers the potential to extend existing location privacy protection capabilities. First, the aim of obfuscation is to protect information about a person's location, but enable that person's true identity to be revealed (thereby avoiding the difficulties faced by anonymity-based approaches, including problems with authentication and personalization). Second, obfuscation does not rely on any centralized server to broker location-based services or administer privacy policies, making it suitable for highly distributed environments like peer-to-peer systems.

Recent work has extended conventional privacy protection strategies using concepts closely related to obfuscation. Gruteser and Grunwald have investigated an anonymity approach called "spatial cloaking" [6]. Similarly, Snekkenes suggest a privacy policy system based on the "need-to-know principle" [14]. However, the work presented in [2] is the first to directly develop obfuscation as a mechanism for protecting location privacy. Obfuscation is a new direction for privacy research that is explicitly spatial and is *complementary* to conventional privacy protection strategies. In contrast to previous work, obfuscation is not based on, but may be used in combination with, regulation, privacy policies, and anonymity.

3 Obfuscation and Negotiation

The aim of obfuscation is to protect a person's location privacy by degrading the quality of information about that person's location, at the same time as delivering a location-based service of acceptable quality to that person. One way to degrade the quality of information about a person's location is to be imprecise. Instead of providing a single location to a location-based service provider, a person might wish to provide a *set* of locations (an *obfuscation set*, usually denoted O). An orthogonal way to degrade the quality of information about a person's location is to be inaccurate. For inaccuracy, we might generate an obfuscation set O that does not contain that person's true location. Thus, we may identify four possibilities for an individual located at point l and their obfuscation set O:

1. Accurate and precise: $l \in O$ and $|O| = 1$
2. Inaccurate and precise: $l \notin O$ and $|O| = 1$
3. Accurate and imprecise: $l \in O$ and $|O| > 1$
4. Inaccurate and imprecise: $l \notin O$ and $|O| > 1$

Under imprecision, the larger the obfuscation set, the less information is being revealed about the individual's true location, and so the greater the level of privacy that individual is able to enjoy. Under inaccuracy, the greater the distance between elements in the obfuscation set and the individual's true location, the less information is being revealed about the individual's true location, and so the greater the level of privacy that individual is able to enjoy.

3.1 Negotiation

Previous work in [2] has provided the formal basis for using obfuscation based on imprecision in nearest POI queries. The approach uses a negotiation process, summarized in algorithm 1 below. The aim of this negotiation process is to achieve a satisfactory balance of level of privacy and quality of service. In this section, we briefly review the negotiation process, but for more details the reader is directed to [2].

Using a graph-based representation of the geographic environment, the negotiation algorithm first partitions the obfuscation set O into equivalence classes. Elements in each equivalence class has the same POI $p \in P$ as their closest (we assume for simplicity there are no ties: all locations $o \in O$ are closest to one POI $p \in P$). This step can be thought of as building a graph-based equivalent of a Voronoi diagram. Indeed, the term "graph Voronoi diagram" is coined and formally defined in [5].

The algorithm operates using a graph-based representation of geographic space in order to model the constraints to movement that are normally a feature of most location-based services. However, the simplest way to explain the negotiation process is with the analogy of a Voronoi diagram. In Figure 1 the Voronoi diagram has been computed for the locations of a few POIs (white dots), with the shaded circular regions in each sub-figure representing the obfuscation set for the clients actual location (black dot). At each iteration of the negotiation process, there are four possibilities:

Algorithm 1. Negotiation proximity query with obfuscation (after [2])

```
Data: Obfuscation set $O$ for the agent; graph-based representation of the geographic
      environment $G$; the set of POIs $P$
Result: The location $p \in P$ which is the best estimate of the nearest POI given the agent's
        privacy requirements
1.1  Find the relation $\delta$ such that for all $o_1, o_2 \in O$, $o_1 \delta o_2$ iff $o_1$ and $o_2$ have the same POI
     $p \in P$ as their most proximal;
1.2  Construct the partition $O/\delta$;
1.3  if $O \in O/\delta$ then
1.4      Return the closest POI for an arbitrary element in $O$;
1.5  if Agent agrees to identify for its current location $l$ the equivalence class $[l] \in O/\delta$ then
1.6      Return the closest POI for an arbitrary element in $[l]$;
1.7  if Agent agrees to identify some new obfuscation set $O'$ such that $O' \subset O$ then
1.8      Reiterate algorithm with $O'$ in place of $O$;
1.9  else
1.10     Return some best estimate of the closest POI based on maximizing $|[l']|/|O|$, for
         some arbitrary $l' \in O$;
```

1. If all the locations in obfuscation set are closest to the same POI (within the same proximal polygon), then the location of this is returned as the query result (Figure 1a, Algorithm 1 lines 1.3–1.4).
2. If the agent agrees to identify in which proximal polygon it is actually location, then the POI for this proximal polygon (shown with bold outline in Figure 1b) is returned as the query result (Algorithm 1 lines 1.5–1.6).
3. If the agent agrees to identify some other smaller obfuscation set (shown as dashed line in Figure 1c), then negotiation reiterates with this new obfuscation set (Algorithm 1 lines 1.7–1.8).
4. Otherwise, the POI for the proximal polygon that contains the largest proportion of the obfuscation set is returned (shown with bold outline in Figure 1d) as a best estimate of the closest POI. Note that, as in Figure 1d, this best estimate may not be the optimal answer (Algorithm 1 lines 1.9–1.10).

The analysis in [2] shows that this negotiation process: (1) is well formed, in the sense that it will always terminate; (2) is computationally efficient, in that its underly-

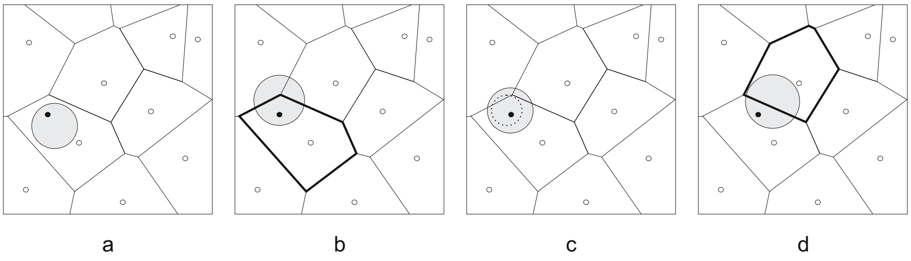

Fig. 1. Negotiation alternatives, illustrated by a Voronoi diagram

ing algorithm has the same computational complexity as the comparable conventional algorithm for finding the most proximal POI without obfuscation.

The decision as to whether an agent agrees to identify the smaller obfuscation sets required for negotiation branches 2 and 3 above will depend on the balance of level of privacy and quality of service for that agent. Thus, if an agent requires a higher quality of service than can be achieved at the current level of privacy, then it may need to reveal more information about its actual location (a smaller obfuscation set). The goal of the remainder of this paper is to investigate this decision, and some of the other parameters that will affect the balance between quality of service and level of privacy for an agent obfuscating its location.

4 Simulations

The framework set out in the previous section provides the basis for an obfuscation system that enables individuals to access high quality location-based services whilst revealing as little information as possible about their current location. Thus, an obfuscation system aims to achieve a satisfactory balance of level of privacy (LOP) and quality of service (QOS). In general, higher LOPs are expected to lead to lower QOS and vice versa (lower LOPs are expected to lead to higher QOS). There exist a number of different parameters that can be manipulated within an obfuscation system and that complicate the balance of LOP and QOS.

In order to investigate these parameters we developed a simulation environment, programmed in Java. The parameters that can be manipulated within the simulation system include:

- the size, shape, and location of the initial obfuscation set used in the negotiation process;
- the strategies adopted by individual agents during the negotiation process;
- the number and location of points of interest (POIs) available within the spatial environment; and
- the spatial environment itself.

By manipulating these parameters within the simulation system we can empirically investigate the effects of these changes upon the balance of LOP and QOS. The primary research questions this research sets out to answer are:

1. Using obfuscation, is it possible to achieve high QOS at the same time as high LOP for location-based proximity queries?
2. Which obfuscation strategies provide the best balance of QOS and LOP?

4.1 Simulation Strategies

There are several distinct obfuscation strategies that have been tested within the simulation system. Each strategy, described below, has different parameters that may be varied as part of the negotiation process.

- *O-strategy*: As described in in the previous section, during the negotiation an agent can choose to identify in which equivalence class it is currently located to the location-based service provider. Using an O-strategy (*optimized*-strategy) agent chooses to reveal this information at the first negotiation iteration. Thus, with minimal negotiation, an O-strategy agent finds the best possible obfuscation for a particular set of initialization conditions. The size of the initial obfuscation is the only negotiation parameter that can be varied by an O-strategy agent.
- *C-strategy*: The C-strategy (*compact*-strategy) uses a full negotiation process that takes advantage of the spatial structure of the environment. The initial obfuscation for a C-strategy agent is a compact connected "ball" of nearby points. At each iteration a C-strategy agent will discard one or more locations from the edge of the obfuscation region, maintaining a compact connected obfuscation throughout the negotiation process. As for the O-strategy, the size of the initial obfuscation is a parameter that can be varied by a C-strategy agent. Additionally, because a C-strategy agent uses a full negotiation process it may also decide to accept a reduced QOS in return for higher LOPs.
- *L-strategy*: The L-strategy (*lying*-strategy) uses inaccuracy rather than imprecision to obfuscate an agent's location. L-strategy agents provide a single (precise) location, but one that is perturbed from the agent's true location. An L-strategy does not take part in any negotiation process, since it provides a precise location from the outset. However, a L-strategy agent can vary the amount it perturbs its true location, i.e., how much it prepared to lie about its true location.

In addition there are a number of derived or hybrid strategies that agents can adopt. Examples of derived and hybrid strategies investigated include the following:

- *E-strategy*: Like the C-strategy, the E-strategy (*elongated*-strategy) uses a connected region of nearby points and maintains a connected obfuscation throughout the negotiation. However, unlike a C-strategy agent, a E-strategy agent uses a elongated "sausage" rather than a compact "ball" of locations for its obfuscation.
- *R-strategy*: Like the C-strategy, the R-strategy (*random*-strategy) uses a full negotiation process but does not take advantage of the spatial structure of the environment. Instead, R-strategy agents construct an initial obfuscation from random locations within the environment. At each iteration, R-strategy agents randomly remove one or more locations from their obfuscation in order to continue the negotiation process.
- *CR-strategy*: CR-strategy agents initialize their obfuscation as a compact region of connected nearby points (C-strategy), but then randomly remove points from that region during the negotiation process.
- *CRL-strategy*: A CRL-strategy is based on a CR-strategy, but additionally an agent may discard its true location from the obfuscation set, meaning it may provide an obfuscation set that is both imprecise and inaccurate.

4.2 Confidence

In addition to the strategies and environmental parameters, some agents may also specify a threshold value which determines how that agent wishes to balance its LOP against its QOS. This threshold value takes the form of a level of confidence, as a number in the interval $[0.0, 1.0]$. At each iteration of the negotiation process, the negotiation algorithm checks the proportion of the obfuscation set that is closest to each POI. If this proportion is greater or equal to the confidence threshold for any of the candidate POIs the agent terminates the negotiation by requesting the best estimate of the nearest POI. A confidence level of 0.6 means that an agent will accept the best estimate of the closest POI as long as 60% of its obfuscation set is closest to one POI. A confidence level of 1.0 means that an agent will only accept a perfect QOS while a confidence level of 0.0 means that an agent will accept any QOS (see [2] for more details). In effect, the confidence level provides a mechanism to balance QOS and LOP, without needing to explicitly compute QOS (which would require that the agent reveal its true location).

5 Results

5.1 Density of Points of Interest

The density of POIs in the environment is one of the main factors that should determine the balance of LOP and QOS. Higher POI densities are expected to require that an agent must reveal more information about its location (lowering its LOP) in order to achieve the same QOS. To investigate this expectation, 100 simulations were conducted at each of 10 different POI densities. The simulations were conducted using a simple environment of a small regular network of 400 nodes arranged in a grid. For each of the 10 sets of simulations, Figure 2 shows the average LOP, in terms of the number of elements in the final obfuscation $|O|$ (i.e., once negotiation is completed), plotted against the against POI density, measured as the number of POIs used to initialize the set of simulations (for a fixed environment size). The median is preferred to the mean as an average, since the population of results for each simulation were often skewed and contained outliers. For clarity, Figure 2 is presented as a log-log plot, because successive sets of simulations doubled the number of POIs in the environment. The other simulation variables were set at default levels: a confidence level of 1.0 was used for all agents (perfect QOS), and each agent used the entire environment as its initial obfuscation (the largest possible initial obfuscation).

The results show that, as expected, LOP decreases with increasing POI density. At the extreme right of the figure, every node is a POI, so every agent must reveal its precise location in order to find the nearest POI with total confidence. At the extreme left of the figure, the environment contains only one POI, so an agent need not reveal any information about its location in order to find the nearest POI (there is only one).

In between these extremes, the figure shows four response curves for the different obfuscation strategies tested. The first strategy is the O-strategy (optimized strategy), where the agent agrees at the first iteration of negotiation to reveal in which equivalence class it is currently located. As expected this strategy outperforms all other strategies, in the sense that it provides a higher LOP than any other strategy for a particular POI density. The worst strategy is the R-strategy (random), where the obfuscation is composed of points located randomly throughout the environment. The poor performance of the R-strategy is is due to the spatially dispersed nature of its obfuscation. Even for small obfuscations, elements of the obfuscation set O may be scattered across the

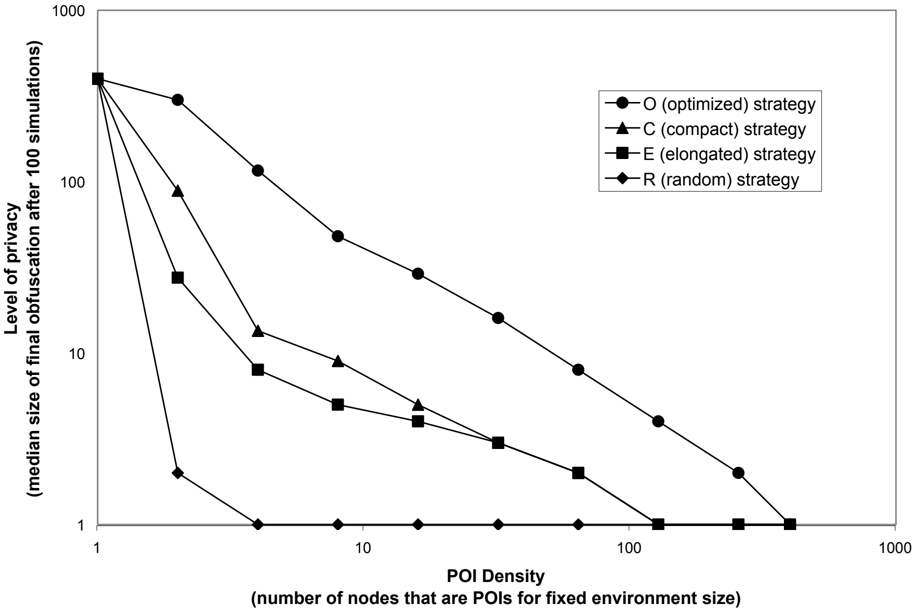

Fig. 2. Effect of POI density upon LOP

environment with no single POI nearest to all of these elements. For similar reasons, the C-strategy, where obfuscations are compact connected subgraphs, outperforms the E-strategy, where obfuscations are elongated connected subgraphs.

A Wilcoxon signed-rank test (a discrete, paired equivalent of the t-test) was used to test the null hypothesis that the observed differences between the results arrived by chance (i.e., sets of results were drawn from the same population). For those data points that are visually distinct on the graph in Figure 2, these tests indicated that the null hypothesis should be rejected at the 5% significance level. In other words, the results indicate that the O-strategy performed as well as or significantly better than the C-strategy, which performed as well as or significantly better than the E-strategy, which performed as well as or significantly better than the R-strategy.

5.2 Initial Obfuscation Size

Another factor that should affect the balance of LOP and QOS is the initial obfuscation size. The size of the initial obfuscation set O constrains the LOP: small obfuscation sets mean lower LOPs, large obfuscation sets allow higher LOPs. To investigate this, 9 sets of 100 simulation runs were performed, changing the size of the initial obfuscation for each set of simulations. Based on the results from Figure 2 the POI density for these simulations was set at a level typical of the mid-range of the simulations (8 POIs in the environment of 400 nodes). The confidence level used in the negotiation process was again 1.0. Figure 3 shows average (median) LOP against initial obfuscation size in terms of $|O|$, the total number of elements in the initial obfuscation set. At the extreme left of the figure, the initial obfuscation size is a single location, leading to the lowest possible LOP. At the extreme right of the figure, the obfuscation size is the entire environment (the agent begins the negotiation process by revealing no information about where it is located).

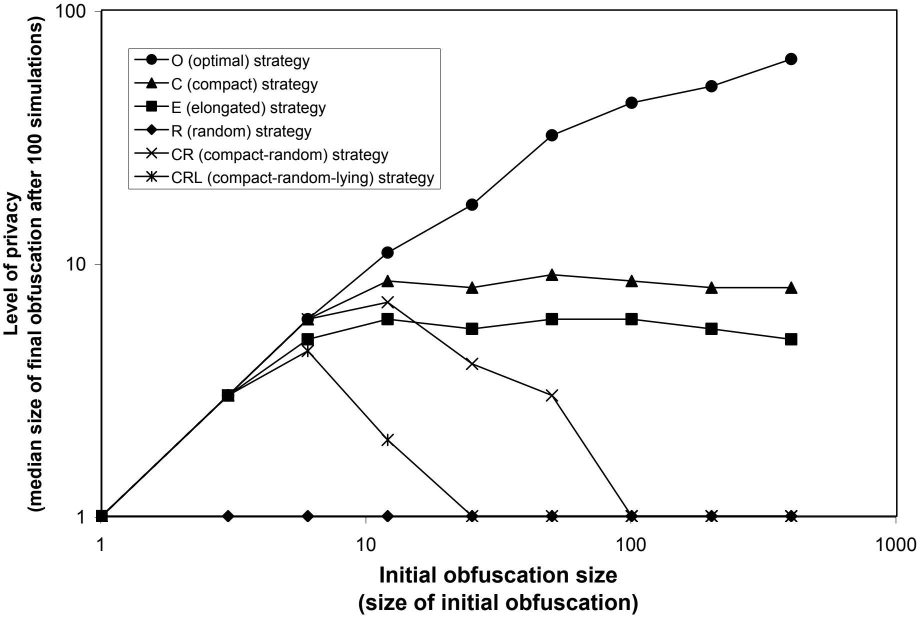

Fig. 3. Effect of initial obfuscation size upon LOP

The figure shows the same ordering of strategies as in the previous section, with the O-strategy outperforming C-strategy, outperforming the E-strategy, outperforming the R-strategy. Again, a Wilcoxon signed-rank test was used to confirm this ordering. Two additional strategies are included in Figure 3, which are hybrids of the C- and R-strategies. The CR-strategy agent selects a compact initial obfuscation in combination with a random negotiation strategy. The CRL-strategy uses the CR-strategy in addition to an agent being able to lie about its true location (i.e., provide an obfuscation that does not include its actual location). These hybrid strategies did not perform well, offering similar performance to the C-strategy only at the lowest initial obfuscation sizes, but degrading rapidly into the R-strategy. In fact, in general these strategies rarely out-performed even the random strategy, providing a strong indication that it is the negotiation process that dominates the effectiveness of the strategy, rather than the initial conditions for the negotiation. This is a helpful result, as is ensures the obfuscation process is not too sensitive to the initial conditions.

The noticeable feature of Figure 3 is that while the LOPs achieved by the C- and E-strategies climb steadily with the O-strategy at lower initial obfuscation sizes, both

C- and E-strategies level off at a maximum LOP at higher initial obfuscation. The maximum level is directly related to the POI density in Figure 2. Repeating these sets of simulations at different POI densities produces similar graphs to Figure 3 which differ in the maxima for the C- and E-strategies. For these sets of related graphs, plotting the maxima for the C-strategy against POI density results in Figure 4.

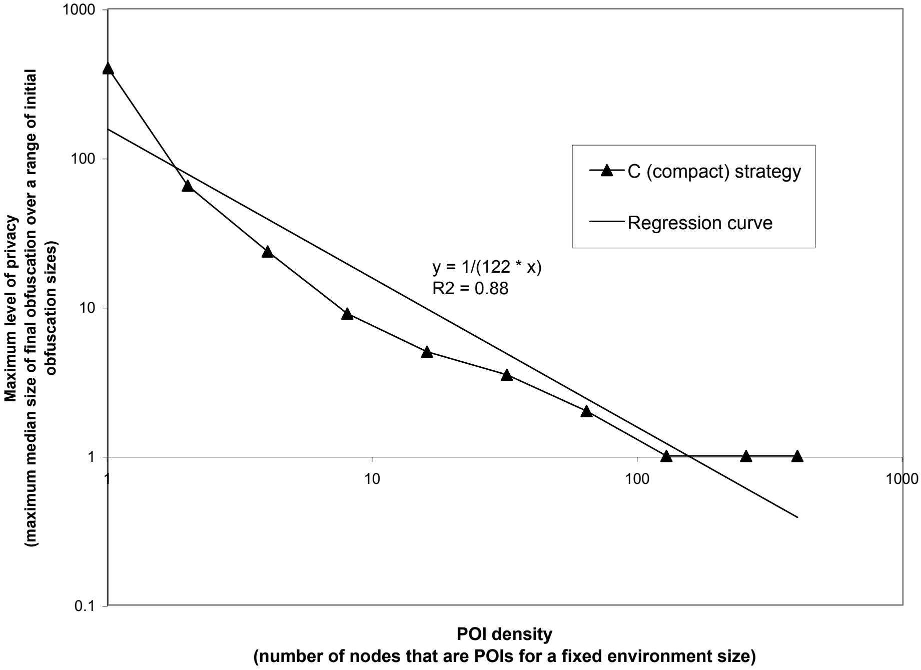

Fig. 4. Maximum median LOP as a function of POI density for a range of initial obfuscation sizes

The import of Figure 4 is that the maximum possible LOP for the C-strategy (and E-strategy) can be directly related to the POI density, independently of the initial obfuscation size. The regression curve in Figure 4 is a simple power function that fits the observed data with a reasonably high product-moment correlation coefficient (r-squared value of 0.88). This is a useful result because knowing that, on average, an initial obfuscation size greater than some threshold will not produce improvements LOP provides a basis upon which to choose an initial obfuscation, one of the primary difficulties facing an obfuscation system.

The reason for this behavior can be explained with the analogy of the Voronoi diagram. For the C- and E-strategies, larger initial obfuscation sets also have a greater overlap with the proximal polygons of several POIs. During negotiation, the obfuscation sets are reduced from the outer boundary. Larger initial obfuscation sets require additional iterations of the negotiation process to ensure that all elements of an agent's obfuscation set are closer to one POI than to any other POIs. Thus, on average the size of the final obfuscation sets are dominated by the density of POIs.

5.3 Confidence

The simulations so far have all used a confidence level of 1.0: the agents need to know with complete confidence which is the nearest POI. While this may be useful in some applications, in many applications users might be prepared to accept suboptimal query results if this also provided higher levels of privacy. Decreasing the level of confidence is expected to increase the LOP, at the expense of decreased QOS. Figure 5 shows the LOPs achieved by the different negotiation strategies across a range of confidence levels. The figure was generated using 11 sets of 100 simulations with a constant mid-range POI density (8 POIs in an environment of 400 nodes) and with the maximum initial obfuscation size.

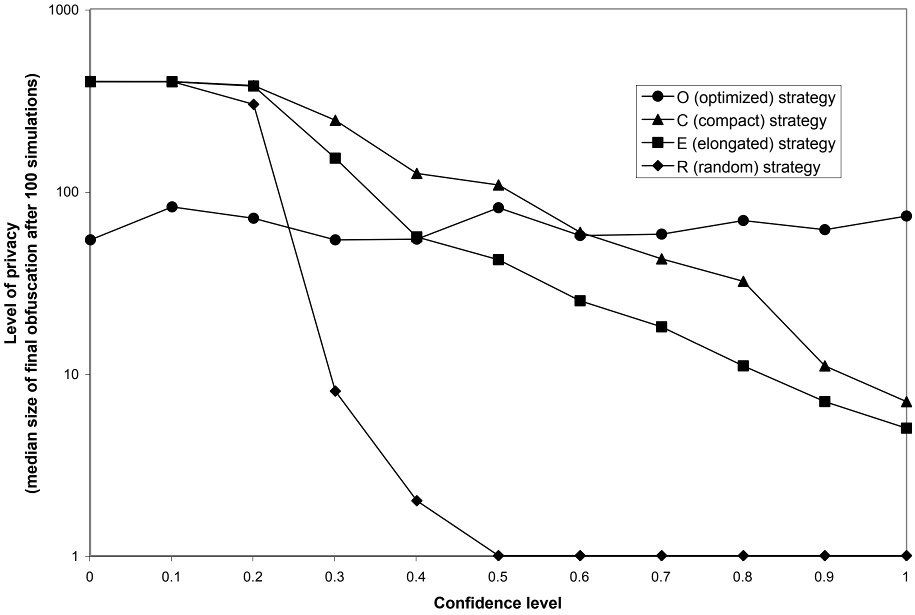

Fig. 5. Effect of confidence level on LOP

As expected, the LOP achieved by the obfuscation system climbs steadily with decreasing confidence levels. Again, the C-strategy outperforms the E-strategy, which in turn outperforms the R-strategy. Both C- and E-strategies still provide moderate levels of privacy even at the highest confidence levels. However, the O-strategy always terminates after one iteration of the negotiation by indicating in which equivalence class it is located. Thus changing the confidence levels has no effect on the performance of the O-strategy. The important point to note in Figure 5 is that at lower levels of confidence, the C-, E-, and even R-strategies are capable of providing higher LOPs that the O-strategy. Thus, for agents who do not require a perfect answer to their location-based query may be able to achieve higher LOPs be using a full negotiation strategy, such as the C-strategy.

A similar set of simulations was used to generate Figure 6, which shows the QOS provided to the agent plotted against level of confidence. QOS is measured in terms of the how much further away from the agent's location a is the actual query result provided by the obfuscation service b when compared with than the best possible query result c measured as $d(a, b) - d(a, c)$ where $d(x, y)$ is the network distance between x and y. High values indicate a low QOS, low values indicate a high QOS (hence the scale in Figure 6 is reversed with low values and high QOS at the top of the y-axis). Although, as expected, QOS decreases with level of confidence, it is noticeable that there exists a wide range of confidence levels less than 1.0, which still provide the highest possible QOS. By definition, the O-strategy always delivers a perfect QOS (although discussed above, sometimes at the cost of lower levels of privacy).

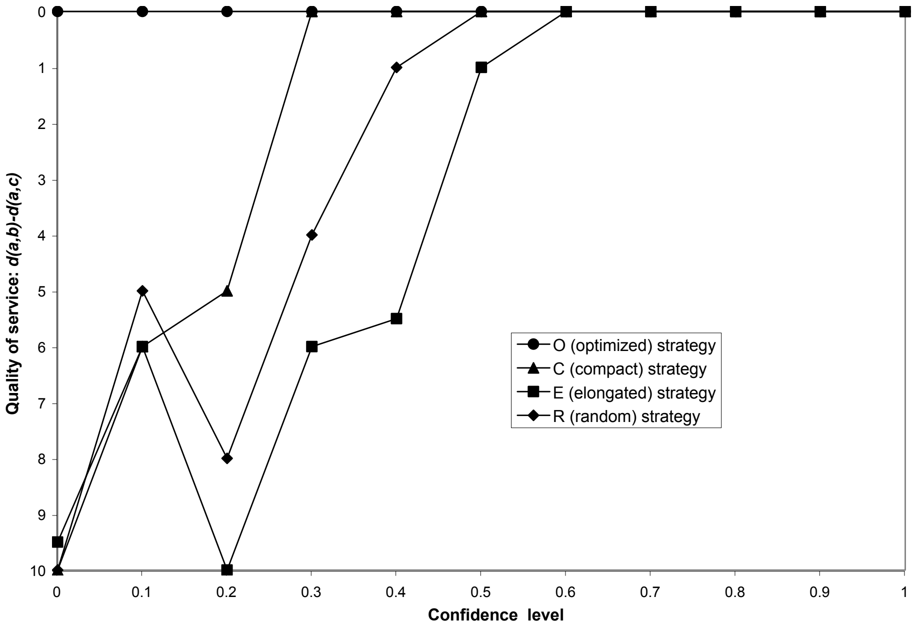

Fig. 6. Effect of confidence level on QOS

5.4 Balancing Quality of Service and Level of Privacy

The previous simulations provide a basic understanding of the effects of varying different obfuscation parameters: different negotiation strategies, POI densities, initial obfuscation sizes, and levels of confidence. We are now in a position to investigate the overall balance of LOP and QOS. Plotting median LOP against median QOS across multiple simulations confirms that it is indeed possible to achieve high QOS and high LOP. Rather than examples of such plots, Figure 7 shows the *lowest* observed QOS against LOP across the entire range of different confidence levels. Thus, Figure 7 represents not the average results, but the worst observed results, in terms of the lowest QOS observed for a particular LOP. Despite representing the worst extremes of the each strategy, the C-strategy in particular still performs well, providing high levels of QOS at all but the highest LOPs. For clarity, the R-strategy performs is not plotted in Figure 7, but as expected performs badly, providing a much lower QOS across almost all LOPs than the other strategies.

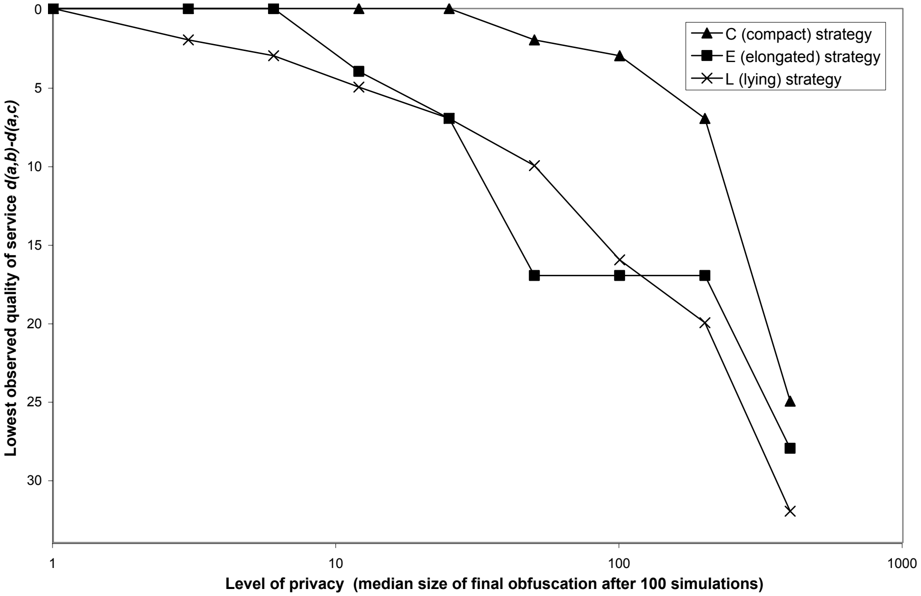

Fig. 7. Lowest observed QOS against LOP

Figure 7 also compares the results with the corresponding L-strategy (lying strategy). The L-strategy provides the LBS with a precise location, but perturbs that location by a predetermined amount. The higher the perturbation, the higher the LOP. To plot the L-strategy and the imprecision-based strategies on the same graph it was necessary to formulate a shared measure of LOP. To achieve this, the LOP for an L-strategy was measured as the size (in terms of the number of elements) of the corresponding C-strategy compact "ball" that has the agent's true location at its center and the perturbed location at its boundary. Based on this relationship, QOS against LOP is also plotted on Figure 7 alongside the C- and E-strategies. Although the L-strategy performs reasonably well, in general the results indicate that the L-strategy does not achieve as high QOS as the C-strategy at the same LOP.

5.5 Environment

As stated previously, the environment for the simulations was a small regular network of 400 nodes arranged in a grid. However, all the experiments were also repeated for a more realistic environment: a generalized map of a portion of central Paris containing a similar number of nodes. The results indicated that, at least within the limited scope of the experiments already described, the use of a less regular environment did not affect the properties of the obfuscation process. However, future work will need to test further types of obfuscation strategy. In particular, the C- and E-strategies tested in this paper use *connected* initial obfuscation sets. A more realistic initial obfuscation set might be all locations within, say, 300m of the agent's actual location. Although in a regular grid environment such an obfuscation will also be connected, in a more realistic, heterogeneous environment, such as central Paris, such an obfuscation will necessarily be connected, which may degrade the performance of the obfuscation process.

5.6 Summary of Results

In summary we draw the following general conclusions from the results of our simulations:

1. Using obfuscation and negotiation, it is possible to achieve high quality location-based services whilst maintaining high levels of location privacy, by revealing only low quality information about location.
2. The final LOP for an agent depends more strongly on the negotiation strategy than on the initial negotiation conditions.
3. Negotiation strategies differ in their suitability for obfuscation. Of the strategies tested, the results suggest that the C-strategy, using a compact region of imprecision and discarding points near the edge of that region during negotiation, outperforms all other non-optimized strategies including a simple L-strategy. Further, at lower confidence levels, the C-strategy can also outperform the O-strategy.
4. The maximum possible LOP for an agent using full negotiation (the C- or E-strategies) across a range of initial negotiation conditions depends strongly on POI density. Within our simulations, the maximum LOP may be related using a simple power law to the POI density. In turn, this provides a maximum initial obfuscation size, above which further increases in initial obfuscation are unlikely to result in concomitant increases in the final LOP.

6 Conclusions and Outlook

Obfuscation enables people to protect their location privacy by revealing the least possible information about their current location when accessing location-based services. Our approach to obfuscation focuses on automated negotiation that enables users to balance the level of location privacy against the quality of location-based service. Since negotiation-based obfuscation is complementary to current approaches to location privacy (regulation, privacy policies, or anonymity), we believe that obfuscation represents an important new direction for location privacy research that has not previously been adequately investigated.

6.1 Balancing QOS and LOP

The simulation results show that a negotiation-based obfuscation strategy for protecting location privacy is able to achieve both high levels of QOS for nearest POI queries in concert with high levels of location privacy. Tailoring the negotiation process to the requirements of a particular user can be achieved using confidence levels. Thus, higher QOS can be guaranteed using high confidence levels; lower confidence levels improves LOP with little or no loss of QOS, at least for higher confidence levels. In summary, these results clearly show that the negotiation process can be used to balance LOP and QOS. These encouraging results warrant further investigation into obfuscation-based strategies, including field tests with real location-based services.

It is expected that in a practical obfuscation system user profiles would be used to govern the automated negotiation process. For example, a simple user profile might contain the minimum LOP a user is prepared to accept (in terms of the minimum obfuscation set size), or the minimum QOS. Current work is investigating the practical aspects of delivering obfuscated location-based services to mobile users.

6.2 Imprecision and Lying

The simulations also indicate that the strategy based on inaccuracy (L-strategy) does not perform as well as the best strategy based on imprecision (C-strategy). This result relies on the establishment of a common measure of LOP for inaccurate and imprecise obfuscation, and must be carefully interpreted. However, it does suggest that the use of imprecision as a core obfuscation strategy warrants further investigation. Imprecision also has the advantage that it does not require the user to explicitly "lie" about his or her location, something that might not be acceptable to some LBS providers. The shape of the obfuscation region has a significant impact for location privacy: compact regions compare favorably with elongated regions. Therefore, strategies that obfuscate a location using regions or blocks should be preferred to strategies that use elongated or linear structures for obfuscation.

6.3 Selecting the Initial Obfuscation Set

Selecting a good initial size for an obfuscation set, which balances location privacy and QOS, is a difficult task. However, our experiments demonstrated that increasing the size of the initial obfuscation set leads to higher levels of privacy only up to a point. These findings suggest that it may be possible to select *a priori* an initial obfuscation region that satisfies the requirements for high LOP and high QOS. This result might be practically useful in a number of ways. For example, a more advanced obfuscation system could enable an obfuscating agent to execute a *prequery*, which determines the average density of POIs within a region. This would only require an agent to reveal minimal information about its current location, but might allow the client to "calibrate" the obfuscation system to select an appropriate initial obfuscation set size that matches the density of POIs.

Other potential mechanisms for setting the initial obfuscation level include: using the entire graph as the initial obfuscation set (computationally practical, as shown in [2]), or selecting "natural" imprecise regions, such "downtown," "Kensington," "Victoria." Such approaches might also be extended to enable obfuscation based on vagueness (e.g., where "downtown" does not have a crisp boundary). Further research on obfuscation for location privacy will address:

- extending obfuscation techniques to other location-based services, in particular navigation services and route queries. Initial work in [3] has already set out the foundations for navigation under imprecision, based on the inherent *instruction equivalence* of navigation instructions.
- extending the static obfuscation model presented in this paper to a truly dynamic model that enables spatiotemporal location-based services.
- counter-strategies for invading location privacy from the perspective of an external agent wanting to undermine a person's location privacy.

Acknowledgments

Dr Duckham is partially supported by an Early Career Researcher Grant and an International Collaborative Research Grant from the University of Melbourne. Dr Kulik is partially supported by an Early Career Researcher Grant from the University of Melbourne. The authors are grateful to Shubham Gupta, who generated the data set of Paris for the simulations, and to the three anonymous reviewers for their helpful comments.

References

1. A.R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
2. M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. In *Pervasive 2005*, Lecture Notes in Computer Science. Springer, Berlin, 2005.
3. M. Duckham, L. Kulik, and M. F. Worboys. Imprecise navigation. *GeoInformatica*, 7(2):79–94, 2003.
4. S. Duri, M. Gruteser, X. Liu, P. Moskowitz, R. Perez, M. Singh, and J-M. Tang. Framework for security and privacy in automotive telematics. In *Proc. 2nd International Workshop on Mobile Commerce*, pages 25–32. ACM Press, 2002.
5. M. Erwig. The graph Voronoi diagram with applications. *Networks*, 36(3):156–163, 2000.
6. M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. MobiSys '03*, pages 31–42, 2003.
7. M. Gruteser and D. Grunwald. A methodological assessment of location privacy risks in wireless hotspot networks. In D. Hutter, G. Müller, and W. Stephan, editors, *Security in Pervasive Computing*, volume 2802 of *Lecture Notes in Computer Science*, pages 10–24. Springer, 2004.
8. J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proc. 2nd International Conference on Mobile Systems, Applications, and Services*, pages 177–189. ACM Press, 2004.
9. E. Kaasinen. User needs for location-aware mobile services. *Personal and Ubiquitous Computing*, 7(1):70–79, 2003.
10. M. Langheinrich. Privacy by design—principles of privacy-aware ubiquitous systems. In G. D. Abowd, B. Brumitt, and S. Shafer, editors, *Ubicomp 2001: Ubiquitous Computing*, volume 2201 of *Lecture Notes in Computer Science*, pages 273–291. Springer, 2001.
11. R.R. Muntz, T. Barclay, J. Dozier, C. Faloutsos, A.M. Maceachren, J.L. Martin, C.M. Pancake, and M Satyanarayanan. *IT Roadmap to a Geospatial Future.* The National Academies Press, Washington, DC, 2003.
12. A. Pfitzmann and M. Köhntopp. Anonymity, unobservability, and pseudonymity—a proposal for terminology. In H. Federrath, editor, *Designing Privacy Enhancing Technologies*, volume 2009 of *Lecture Notes in Computer Science*, pages 1–9. Springer, 2001.

13. B. Schilit, J. Hong, and M. Gruteser. Wireless location privacy protection. *IEEE Computer*, 36(12):135–137, 2003.
14. E. Snekkenes. Concepts for personal location privacy policies. In *Proc. 3rd ACM conference on Electronic Commerce*, pages 48–57. ACM Press, 2001.
15. M. F. Worboys and E. Clementini. Integration of imperfect spatial information. *Journal of Visual Languages and Computing*, 12:61–80, 2001.