Road Network-aware Anonymization in Mobile Systems with Reciprocity Support

Bhuvan Bamba[♢], Ling Liu[♣], Emre Yigitoglu[♣]
[♢] Nextdoor, Inc.
[♣] College of Computing, Georgia Institute of Technology
bhuvan@nextdoor.com, {lingliu, eyigitoglu}@cc.gatech.edu

***Abstract*— Most existing solutions on location anonymization fail to address the issue of location privacy protection for mobile users traveling on road networks. In this paper, we present a road network-aware privacy model to handle the location privacy problem in mobile systems. We present third party supported anonymization which can be plugged into existing mobile systems without major modifications. As opposed to current road network specific solutions, our system has the following salient features. First, we argue that the graph density of a cloaked subgraph is an important location cloaking quality measure for road network-based location privacy protection. High graph density leads to high utility of the cloaked location with respect to spatial resolution. Second, we devise a suite of graph-based cloaking algorithms which guarantee reciprocity - an important location cloaking property that many existing approaches fail to support - under an enhanced privacy model. Further, we use controlled randomization in the cloaking process to provide higher privacy strength while maintaining the utility of the cloaked location. Last but not the least, our experimental evaluation shows the efficiency and robustness of the approach in terms of privacy, utility and performance of the model.**

I. Introduction

The widespread availability of location sensing technologies has led to the successful deployment of a large number of location-based services and applications. The plethora of services available today provide a huge scope for business opportunities. However, the availability of continuous location information opens the door for potential misuse [11] as it can be used for stalking, performing inference about a user's medical conditions, or for location-based spam.

Location privacy refers to the capability of enabling a mobile node or a trusted server to conceal the relation between location and personal identifiable information from third parties. A fair amount of work has been performed to address the problem of location privacy and most existing research can be classified into spatial cloaking [1], [6], [8], [15], [7] and mix zone-based solutions [2], [9], [4], [5], [18]. Most of the anonymization work focuses on privacy-utility trade-offs of the anonymized location information. However, this fails in both privacy protection and utility preservation for mobile users traveling on road networks.

The main problem arises due to the fact that the attained privacy level falls below the promised privacy level when the underlying road network is considered. As an example, we consider spatial cloaking techniques [1], [6], [8], [15] where the exact user position is anonymized to an area containing other users, thus rendering an adversary incapable of identifying the actual user position within the cloaked location. However, when the underlying road network is considered it may be possible for the adversary to identify that the user is moving on a particular road segment. This unauthorized exposure of segment-identity relationship may have undesirable consequences. For example, if the road segment can be identified as the only road to enter a religious or a political club, then the probability of association of the mobile user with this sensitive public location can be intrusive in terms of privacy.

XStar [20] presents the first work of combining segment l-diversity with location k-anonymity using star-based cloaking, aiming at offering privacy-aware mobile services over road networks. Concretely, by segment l-diversity definition introduced in [20], a subgraph of l connected segments without a high degree forking junction can be a cloaked subgraph. Such subgraphs often fail to preserve the true l-segment diversity since an adversary can easily link a request with a road segment in such a simple line graph of l segments with higher probability than $1/l$. An extreme case is when these l segments are small logical segments of an actual long road segment linking two star structures.

Another problem with XStar [20] is the lack of explicit support of the reciprocity criterion for successful cloaking. Informally, reciprocity states that by satisfying location k-anonymity, there should be at least k users using the same cloaked location in their service requests. However, many existing works [8], [15], [1] fail to meet the reciprocity requirement since their definition of location k-anonymity simply requires that there are $k-1$ other users residing in the same cloaked location. As a result, instead of cloaking k users with a unique cloaking region, each of the k users is cloaked separately, resulting in different cloaking regions being reported by each of these k users. The probability of linking a user with a request is much higher than $1/k$ when reciprocity is not met [7].

In this paper, we present MobiCloak, a reciprocity preserving road network-aware location privacy model for users traveling on road networks. To address the problems in existing solutions, MobiCloak defines segment s-anonymity as a companion metric to the user k-anonymity metric. Our k-anonymity definition explicitly addresses the reciprocity requirement whereas segment s-anonymity mandates forking

junctions as a necessary constraint to strengthen privacy measures. The MOBICLOAK approach makes three unique contributions. First, we argue that the graph density of a cloaked subgraph is an important quality measure for road network-based location privacy protection. High cloaking graph density leads to high utility in terms of spatial resolution in the road network. Compact subgraphs lead to lower query processing costs based on current road network-based query evaluation cost models [19].

Second, we present graph-based cloaking algorithms, offering different levels of optimization in terms of privacy-utility trade-offs. Concretely, we use *randomized expansion* and *network expansion* as two naïve baseline techniques, each of which represents one end of the privacy-utility spectrum. As opposed to the work presented in [20], our algorithms ensure reciprocity of requests which are anonymized together. The randomized expansion technique, although it displays high attack resilience, suffers from poor utility. In order to deal with the static nature of the network expansion technique, which leads to low attack resilience, we introduce service request density-aware techniques, which utilize additional controlled randomness to provide higher resilience to replay attacks, while maintaining desired graph density of cloaked location. Last but not the least, we evaluate the performance of MOBICLOAK anonymization algorithms through experimental evaluation.

The rest of this paper is organized as follows. Section II details the road network-aware privacy model used in MOBICLOAK, followed by an introduction to our graph-based location anonymization algorithms in Section III. Section IV reports our experimental evaluation of the MOBICLOAK algorithms using a real world road network-based simulator. The paper provides an overview of the related work and concludes with a summary in Section V.

II. NETWORK-AWARE PRIVACY MODEL

In this section lays the groundwork for our road network-aware location cloaking algorithm development. We formally introduce the MOBICLOAK location privacy model, the anonymization procedure, and the evaluation metrics to study the effectiveness of our approach.

A. MobiCloak Location Privacy Model

In MOBICLOAK, each mobile user has a personalized privacy preference profile (P3P) [1]. Two measures are used to capture the user-level location privacy requirements: user k-anonymity and segment s-anonymity. The first requirement ensures that the probability of an adversary being able to link a mobile user with a request generated from the cloaked subgraph is no greater than $1/k$. The companion requirement of segment s-anonymity ensures that the probability of an adversary being able to link a request with any road segment in the cloaked subgraph is no greater than $1/s$. This addresses the vulnerability of associating a user with a particular segment, which can be used to link the user to a sensitive public location such as some religious club or AIDS treatment center.

User k-anonymity cannot be guaranteed without ensuring segment s-anonymity in the road network-based model. For example, in Figure 1 a request from user u_1 with privacy parameters $k = 5$ and $s = 3$ is anonymized to include segments r_3, r_4 and r_5. As shown by the dotted bounding region, five users $\{u_i\}_{i=1}^{5}$ are present within the region on the road segments. As long as the adversary is unable to associate a road segment with the service request, the probability of associating any user with the request is no greater than $1/5$. However, if the adversary is able to associate the request with a particular segment, say r_3 in the figure, the user k-anonymity effectively falls to two. With these observations in mind, we argue that an anonymized request must satisfy the dual requirements of user anonymity and segment anonymity.

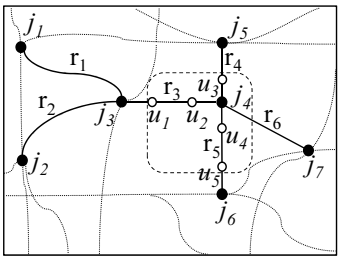


Fig. 1: Road Network-based Model

Definition 1: (User k-Anonymity) A user location is said to be k-anonymous if there are $(k-1)$ other users within the same location released by the anonymization server.

Definition 2: (Segment s-Anonymity) The cloaked location of a user request is said to be segment s-anonymous if it contains at least s distinct road segments, including the road segment associated with the user's actual location, as well as at least one forking junction of degree higher than two.

Our definition of segment s-anonymity removes the possibility of producing a successful cloaked line subgraph. To further strengthen the attack resilience, we incorporate graph-density metric in both our cloaking algorithms for selecting the best candidate segment to expand and the evaluation of MOBICLOAK.

B. Location Anonymization Procedure

MOBICLOAK can be used by a trusted third party anonymizer to mediate between mobile users and LBS providers. We assume that given the current location of a mobile user, the anonymization server can determine the road segment associated with any generated service requests based on the underlying road network topology. All service requests issued by mobile users are forwarded to the anonymization server and represented by a message of the form $m_s = \langle obj_{id}, req_{id}, \{r, t\}, k, s, \sigma_s, \sigma_t \rangle$. The first two components of the message $-$ obj_{id} and req_{id} uniquely identify a request. The spatio-temporal location of the message m_s is represented by the road segment associated with the current user location r and the timestamp t. The four parameters $\{k, s, \sigma_s, \sigma_t\}$ denote the P3P for the user issuing the service request. σ_s denotes the spatial tolerance of the request which affects the size of the candidate result set [1] and σ_t denotes the service delay which the user is willing to accept; together they constitute the utility metrics of the service request.

The general procedure of location anonymization is to take a service request m_s as input and apply a specific road network-aware cloaking algorithm to generate a perturbed message $m_t = \langle h(obj_{id}||req_{id}), \mathcal{S}\rangle$, where h is a secure hash function, $\mathcal{S}$ denotes the cloaked subgraph comprising at least s road segments, including r. Additionally, the cloaked subgraph $\mathcal{S}$ must meet the *reciprocity* requirement. We formally define the reciprocity criterion using the notation of MOBICLOAK privacy model.

Definition 3: (Reciprocity) A cloaked subgraph $\mathcal{S}$ is said to meet the reciprocity requirement, iff $\mathcal{S}$ contains a set of service requests M such that (i) $|M| \geq max_{i=1}^{|M|}\ m_i.k$, (ii) $|\mathcal{S}| \geq max_{i=1}^{|M|}\ m_i.s$, (iii) $\forall\ m_i \in M$, $\mathcal{S}$ is a subgraph of $B_{max}(m_i.\sigma_s, m_i.\sigma_t)$.

The cloaked subgraph $\mathcal{S}$ lies within the *maximal spatio-temporal cloaking box* $B_{max}(m_i.\sigma_s, m_i.\sigma_t)$ of each message $m_i \in M$. Thus, an optimal algorithm would partition the road network graph into *reciprocity conformant* subgraphs which have minimal spatial resolution. However, optimal k-anonymity on its own is an NP-Hard problem [14]. In MOBICLOAK, the key idea is to provide controlled randomness when we construct a cloaking subgraph by expanding from the current segment on which a request resides. Concretely, to facilitate our implementation of reciprocity, we associate a segment profile κ_r with each road segment r as the basic data structure for MOBICLOAK algorithms to make an informed decision of segment expansion at each algorithmic step.

Definition 4: (Segment Profile κ_r) Each road segment r is associated with a segment profile $\kappa_r = \langle M_r, max_{i=1}^{|M_r|}\ m_i.k,\ max_{i=1}^{|M_r|}\ m_i.s,\ min_{i=1}^{|M_r|} m_i.\sigma_s, min_{i=1}^{|M_r|} m_i.\sigma_t\rangle$, where M_r denotes the set of messages associated with r.

C. Evaluation Metrics

In this section, we define three sets of metrics that will be used to evaluate the cloaking algorithms. The first set of metrics is used to evaluate the level of privacy protection, and includes relative k-anonymity (k_{rel}), relative s-anonymity (s_{rel}), and segment entropy ($\mathcal{H}(\mathcal{S})$). The second set of metrics is used for evaluating the level of utility preserved in the cloaked subgraph produced by a cloaking algorithm. Important utility measures include relative spatial resolution ($\sigma_{srel}(\mathcal{S})$), relative temporal resolution ($\sigma_{trel}(\mathcal{S})$) and graph density $\rho(\mathcal{S})$. The third set of metrics, anonymization success rate (R) and anonymization time t, is used for evaluating the cloaking algorithm performance.

Relative k-anonymity (k_{rel}) and Relative s-anonymity (s_{rel}): These two metrics measure the achieved levels of location k-anonymity and segment s-anonymity for successfully cloaked messages. Given a set of messages M cloaked together, we define $k_{rel} = \frac{1}{|M|}\sum_{m_s \in M} \frac{|M|}{m_s.k}$. Similarly, $s_{rel} = \frac{1}{|M|}\sum_{m_s \in M} \frac{|S|}{m_s.s}$. Although our cloaking algorithms aim at obtaining higher anonymity, high anonymity achieved at the cost of a larger cloaked subgraph leads to higher processing cost of service requests.

Segment Entropy ($\mathcal{H}(\mathcal{S})$): The segment entropy is a measure of the privacy achieved by our cloaking techniques. We measure segment entropy as $\mathcal{H}(\mathcal{S}) = -\sum_{i=1}^{|\mathcal{S}|}\ p_i.\log p_i$, where p_i denotes the probability of the user position being associated with the i^{th} segment in $\mathcal{S}$.

Relative Spatial Resolution ($\sigma_{srel}(\mathcal{S})$): This metric measures the ability of a cloaking algorithm to provide compact subgraphs. Concretely, we define the $\sigma_{srel}(\mathcal{S})$ over a set of messages M by $\frac{1}{|M|}\sum_{m_s \in M} \frac{m_s.\sigma_s}{\sigma(\mathcal{S})}$, where $\sigma(\mathcal{S})$ represents the spatial resolution (radius) of the cloaked subgraph $\mathcal{S}$.

Relative Temporal Resolution ($\sigma_{trel}(\mathcal{S})$): This metric measures the ratio of the maximum allowable delay σ_t over the amount of delay introduced by a perturbed message m_t. Let $I = [t_s, t_e]$ denote the time interval between the earliest and latest message $\in M$. We define $\sigma_{trel}(\mathcal{S})$ over a set of messages M by $\frac{1}{|M|}\sum_{m_s \in M} \frac{m_s.\sigma_t}{|m_t.I|}$.

Average Graph Density ($\rho(\mathcal{S})$): The average graph density is a measure of the compactness of the cloaked subgraph. The graph density for a cloaked subgraph is defined as $\rho(\mathcal{S}) = \frac{2.|E|}{|V|.|V-1|}$ where $|V|, |E|$ denote the number of nodes, edges respectively in the cloaked subgraph.

Anonymization Success Rate (R): Let M denote the set of anonymization requests received by the system. The set of messages that are successfully perturbed can be computed by $\{m_t | m_t = g_{cloak}(m_s),\ m_s \in M\}$, where $g_{cloak}(m_s)$ denotes an anonymization algorithm. The success rate is defined as follows: $R(g_{cloak}(m_s)) = \frac{|\{m_t | m_t = g_{cloak}(m_s), m_s \in M\}|}{|M|}$.

Message Anonymization Time (t): This metric measures the run-time performance of the cloaking algorithm in terms of time complexity. Efficient cloaking implies that the cloaking algorithm spends less time to perturb more messages.

III. ANONYMIZATION ALGORITHMS

In this section, we present two sets of graph-based cloaking algorithms. Each algorithm accepts as input a message m_s for a service request. Neighboring segments in the road network are selected starting from the segment associated with m_s. The segment selection process is iterative and each step aims at meeting the privacy constraints k and s without violating the utility constraints σ_s and σ_t for all requests associated with the set of selected segments. The efficiency of a cloaking algorithm is determined by its segment selection heuristic.

A. Naïve Baseline Cloaking Algorithms

The first set of cloaking algorithms serve as naïve baseline methods and includes basic network expansion and randomized expansion. Both algorithms make the segment expansion decision solely based on the underlying road network topology. We now illustrate their inability to provide a good balance between privacy and utility of cloaked locations.

Network Expansion: Network expansion is designed using a variation of Djikstra's shortest path algorithm. The procedure starts by initializing the cloaked subgraph with the segment associated with the initial service request. Segment profile κ_r is used to determine the current privacy and utility constraints. The procedure iterates the segment expansion process by first checking if the current subgraph satisfies the privacy requirements. If not, the algorithm selects the road segment

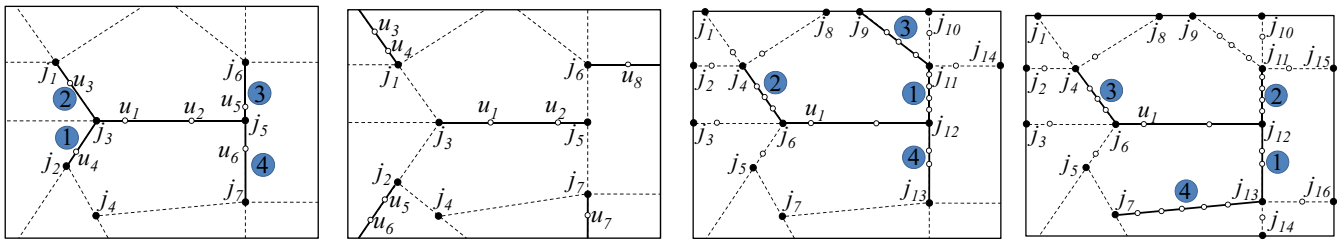


Fig. 2: (a) Network Expansion (b) Randomized Expansion (c) Density-Aware Expansion (d) Density-Aware Randomized Expansion

that is closest to the user's current position in terms of road network travel distance [19] and satisfies the reciprocity requirement. The selected segment is added to the intermediate cloaking subgraph, and anonymization parameters are updated with the profile of the newly added segment. If the current subgraph meets the reciprocity requirements, the algorithm terminates successfully. Otherwise, the iterative procedure continues till the reciprocity requirements are met (success) or no new segment can be added to the current subgraph (failure). The cloaked subgraphs generated are compact with small spatial extent and high graph density. However, this approach suffers from poor attack resilience since it uses a deterministic expansion heuristic. Figure 2(a) displays an example anonymization using this approach with the order of addition of segments denoted by the encircled numbers.

Randomized Expansion: Randomized expansion computes a segment pool comprising of segments with profiles satisfying maximal spatial and temporal reciprocity requirements. Then, starting with the initial segment, it randomly selects a segment from the segment pool to add to the cloaked subgraph. For example, Figure 2(b) shows the same example fragment of the road network where user u_1 issues a request. The profile of segment $\overline{j_3 j_5}$, has $(k, s) = (5, 3)$. However, request from user u_8 has $s = 5$ which modifies the overall privacy requirements. The algorithm terminates when the cloaked subgraph meets the reciprocity criteria for privacy requirements. Randomized expansion allows for disconnected segments, which results in less compact subgraphs, and high service processing costs. However, using a set of disconnected segments as the cloaked location offers high resilience to replay attacks.

In summary, the use of the underlying road network topology in the network expansion algorithm makes the segment selection static and deterministic in nature and thus results in low attack resilience. On the other hand, randomized segment expansion leads to low graph density, though it offers high attack resilience.

B. Service Request Density-Aware Cloaking

In order to overcome the inability of naïve approaches to balance privacy-utility trade-offs, we develop density-aware dynamic cloaking techniques. The main idea is to utilize the segment profiles (mainly $|M_r|$) associated with each segment, which is dynamic in nature, to determine which segment should be added next to the cloaked subgraph. Given that the request count $|M_r|$ is highly dynamic, an adversary with knowledge of the underlying road network is unable to perform replay attacks successfully. We show that density-aware expansion can generate subgraphs with higher graph density; yet it is more resilient than naïve network expansion. Density-aware randomized expansion introduces controlled randomness in the cloaking process in order to strengthen the attack resilience against adversarial background knowledge.

Density-Aware Dynamic Expansion: Density-aware dynamic expansion takes a service request m_s and starts with the initial cloaking subgraph $\mathcal{S}$ containing only the segment r associated with m_s. Two priority queues are created and maintained: segment count-based node priority queue Q_n and request count-based segment priority queue Q_s. The two end nodes of r are used to initialize Q_n and are inserted in the decreasing order of the segment count connected to the node. The algorithm then inserts all segments connected to these nodes which satisfy the spatial and temporal reciprocity requirements into Q_s in decreasing order of their request count. The segment at top of Q_s is added to the cloaked subgraph and the anonymization profile is updated based on the newly added segment. Furthermore, the other end node of this segment is added to Q_n. All segments connected to this node junction, satisfying the reciprocity requirement, are inserted into Q_s. This procedure is repeated till either a cloaked subgraph which meets the reciprocity criteria is found (success) or no new segment can be added (failure).

Consider the example in Figure 2(c) where user u_1 issues a request with $(k, s) = (5, 5)$. For simplicity, we assume that other requests in the neighborhood have lower privacy requirements and do not violate spatial and temporal reciprocity requirements. The segment $\overline{j_6 j_{12}}$ is added to the cloaking subgraph. The algorithm next inserts the segments $\overline{j_3 j_6}$, $\overline{j_4 j_6}$, $\overline{j_5 j_6}$, $\overline{j_{11} j_{12}}$, $\overline{j_{12} j_{13}}$ connected to junctions j_6 and j_{12} into Q_s in the decreasing order of their request count. Segment $\overline{j_{11} j_{12}}$ has the highest request count of four and is selected to add to the cloaking subgraph which currently contains the segment $\overline{j_6 j_{12}}$ only. At this point, the k-anonymity requirements are met but the s-anonymity requirement is not satisfied. Thus the algorithm continues to expand the cloaking subgraph. The procedure is repeated till the cloaking subgraph meets the segment s-anonymity requirement.

This approach leads to much higher k-anonymity ($k = 14$ in the above example) compared to network expansion approach, thus providing higher privacy with similar QoS. Furthermore, unlike network expansion, density-aware expansion uses re-

quest counts, which are dynamic in nature, as an expansion criterion making it harder for an adversary to mount an attack. However, if an adversary could continuously log requests on the road network, it is possible to perform a replay attack.

Density-Aware Dynamic Randomized Expansion: In order to overcome the deficiency of density-aware technique, we design an algorithm that can break the deterministic nature of the request count-based expansion. Introduction of controlled randomness makes the generated subgraphs highly attack resilient.

Concretely, this algorithm requires a modification to the previous approach. While selecting a segment to add to the cloaking subgraph the algorithm does not select the segment at the top of Q_s. It considers all segments with request counts in the range of $[\lceil |M_r^{top}| * (1-\alpha) \rceil, |M_r^{top}|]$ and randomly selects a segment to add to the cloaked subgraph from this subset. $|M_r^{top}|$ represents the request count associated with the segment at the top of Q_s and α represents the randomization factor supplied by the system. This approach makes it impossible for the adversary to launch a replay attack even in the presence of complete knowledge of the segment profiles and service request positions.

We illustrate this algorithm through Figure 2(d), where user u_1 performs a request with $k = 5$ and $s = 5$. The request is anonymized using the randomization factor $\alpha = 0.5$. The request lies on the segment $\overline{j_6 j_{12}}$. The algorithm next inserts segments $\overline{j_3 j_6}$, $\overline{j_4 j_6}$, $\overline{j_5 j_6}$, $\overline{j_{11} j_{12}}$, $\overline{j_{12} j_{13}}$ connected to junctions j_6 and j_{12} into Q_s in order of their request count. Segment $\overline{j_{11} j_{12}}$, with a request count of four, lies at the head of Q_s. All three segments with mobile object counts in the range of [2, 4] are considered while expanding the cloaking subgraph. The algorithm may randomly select any of the three segments to extend the current cloaked subgraph. Figure 2(d) shows the algorithm selects $\overline{j_{12} j_{13}}$ as the next segment and proceeds in an iterative manner. The effect of the randomization factor is studied in the next section.

IV. Experimental Evaluation

In this section, we perform an empirical analysis of the algorithms - Network Expansion (NE), Randomized Expansion (RE), Density-Aware Dynamic Expansion (DAE) and Density-Aware Dynamic Randomized Expansion (DARE) - based on the metrics defined in Section II-C. The experimental evaluation is focused on the effectiveness of our cloaking algorithms in terms of privacy and QoS metrics, entropy-based privacy measure and performance of the anonymization algorithms.

A. Experimental Setup

Our simulator generates a trace of vehicles moving on a real-world road network using maps available from the National Mapping Division of the U.S. Geological Survey. Vehicles are uniformly placed on the road network according to traffic densities determined from the traffic volume data in [8]. The simulator simulates the motion of 20,000 users on the road network with appropriate velocity information, as they follow the shortest path to their destination, generating requests at mean interval of 30 seconds.

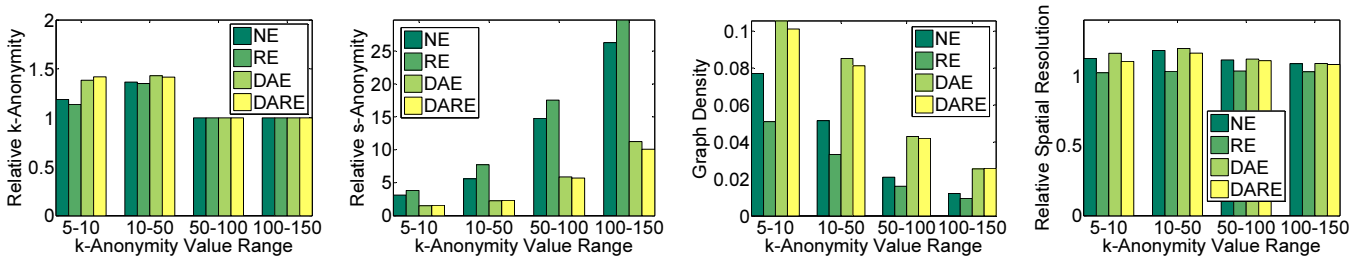
We use maps of varying sizes to observe the performance of the MobiCloak algorithms under different conditions. Results reported in this paper are measured using a map of Chamblee region of Georgia, which covers an area around 168 km^2 in expanse, to generate the trace. The map comprises of about 10,000 road segments and 7,000 road junctions. Our experiments use traces generated by simulating vehicle movement for a period of one hour, results are averaged over a number of such traces. Default k values are chosen from the range $10-50$ and default s values are chosen from the range $10-20$ along a Zipfian distribution with parameter value 0.6. Default values for σ_s and σ_t are $1-1.5$ km and 30 seconds respectively. The default values for the randomization factor α and λ are set to 0.5 and 2 respectively. Unless mentioned otherwise, parameters are set to their default values.

B. Experimental Results

We first present the experimental results for user-defined privacy and QoS metrics, the entropy-based privacy metric, and the system performance by varying the values of k-anonymity. Our results show that (i) the density-aware techniques are more effective than naíve approaches on both privacy and QoS metrics, (ii) density-aware approaches may have slightly lower success rate as they are prone to selecting longer segments thus violating the σ_s requirement in a few cases, (iii) DARE has attack resilience close to the RE approach, (iv) spatio-temporal cloaking allows for trade-offs between spatial and temporal resolution of cloaked messages.

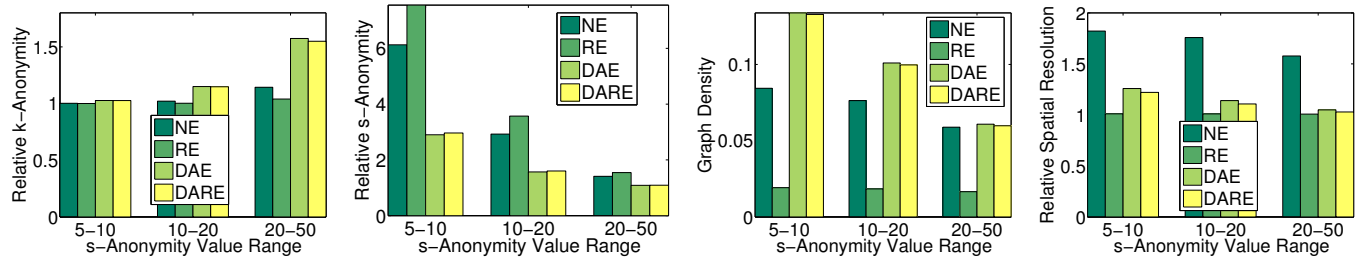
1) User-defined Privacy and QoS Metrics: **Varying k-Anonymity Levels:** The performance of the system is observed as we vary the k-anonymity values to test the performance of our algorithms for varying user anonymity levels. We set the s-anonymity levels between $[5, 10]$, values chosen along a Zipfian distribution with parameter 0.6, maximum spatial resolution values between $[2, 3]km$ and maximum temporal resolution with a mean value of 30 seconds.

Figure 3 measures the relative k, relative s, average graph density, and relative spatial resolution (extent) with varying k values. The k-anonymity levels are chosen along a Zipfian distribution with parameter 0.6 from the following set of range values: $[5, 10]$, $[10, 50]$, $[50, 100]$ and $[110, 150]$. In Figure 3(a), the relative k-anonymity values are higher for the mobility-aware approaches compared to naive baseline randomized expansion and network expansion approaches for lower k-anonymity levels. This is because the mobility-aware approaches select segments with higher number of mobile objects; the mobility-aware randomized approach has only slightly lower relative k-anonymity compared to the mobility-aware approach. For higher k-anonymity requirements, all approaches have relative k-anonymity levels close to one. However, as shown in Figure 3(b), the mobility-aware approaches achieve these k-anonymity levels by selecting fewer number of segments compared to the baseline randomized expansion and network expansion approach. This is visible



(a) Relative k-Anonymity Level (b) Relative s-Anonymity Level (c) Average Graph Density (d) Relative Spatial Resolution

Fig. 3: User-defined Privacy and QoS Metrics with Varying k-Anonymity Levels



(a) Relative k-Anonymity Level (b) Relative s-Anonymity Level (c) Average Graph Density (d) Relative Spatial Resolution

Fig. 4: User-defined Privacy and QoS Metrics with Varying s-Anonymity Levels

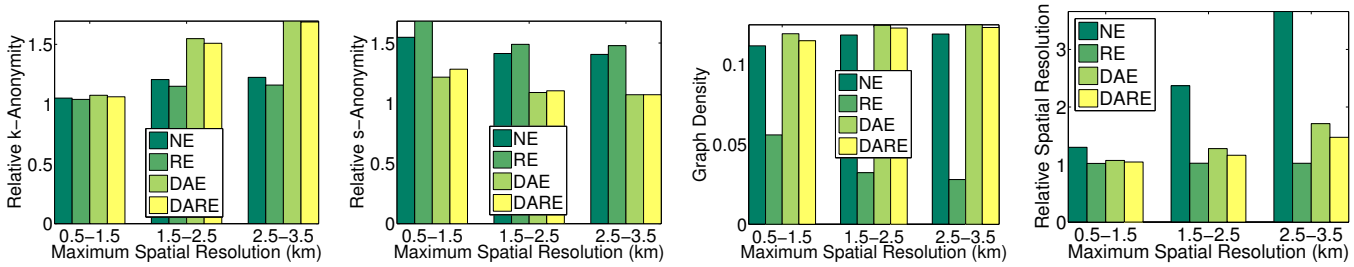
by observing the lower relative s-anonymity levels achieved by the mobility-aware approaches, particularly for higher k values.

Figure 3(c) displays the average graph density achieved by the four cloaking approaches as we vary the k-anonymity requirements. The mobility-aware approaches achieve consistently higher graph density compared to the randomized approach. Another observation is that the cloaking subgraphs produced by the mobility-aware schemes are compact enough to outperform the network expansion approach. As k increases, the graph density achieved by all the approaches falls as larger subgraphs are produced; however, the mobility-aware approaches maintain their advantage over the baseline randomized expansion and network expansion approach. As a final quality measure, we observe the relative spatial resolution (extent) achieved by all four approaches as shown in Figure 3(d). The baseline network expansion approach outperforms all the other approaches on this metric. The mobility-aware approaches perform slightly better than the randomized expansion approach for higher k values. The lower relative spatial resolution of mobility-aware approaches results from the biased selection of segments containing larger number of mobile objects. Longer road segments are expected to have larger number of mobile users traveling on them and are typically listed at the top of the priority queue. These segments are given higher priority for segment-based expansion in constructing the cloaked subgraph.

Varying s-Anonymity Levels: This set of experiments tests the performance of the algorithms for varying s-anonymity levels. We set the k-anonymity level between $[50, 100]$ values chosen along a Zipfian distribution with parameter 0.6, maximum spatial resolution values between $[2, 3]km$ and maximum temporal resolution with a mean value of 30 seconds.

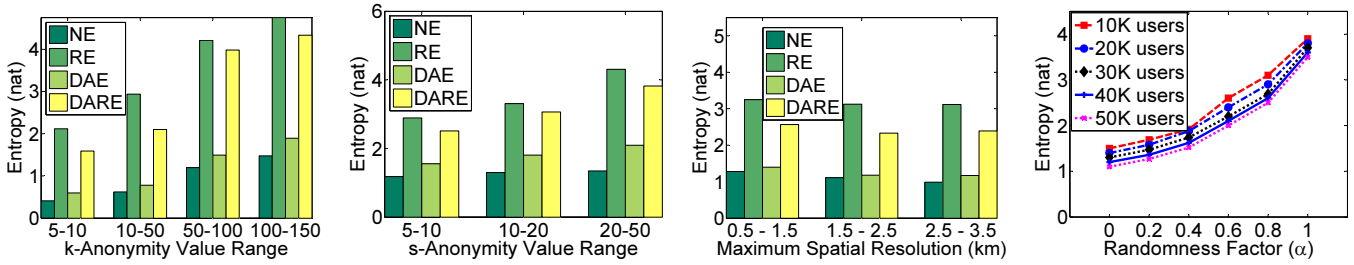
Figure 4 shows the measurements of relative k, relative s, average graph density, and relative spatial resolution with varying s. The s-anonymity levels are chosen along a Zipfian distribution with parameter 0.6 from the following set of range values: $[5, 10]$, $[10, 20]$ and $[20, 50]$. The relative k-anonymity levels increase as we increase the required s-anonymity levels (Figure 4(a)). This is largely due to the fact that a larger number of segments need to be selected for meeting the s-anonymity requirements, which results in larger k values. As expected, our mobility-aware approaches have higher relative k-anonymity levels as they select segments with large number of mobile users. Figure 4(b) displays the relative s-anonymity levels which are high for lower s values. This is due to the fact that larger number of segments need to be selected to meet the required k-anonymity levels. As we increase the required s levels, the relative s-anonymity levels fall. Secondly, the mobility-aware approaches have lower relative s-anonymity levels as they meet the k-anonymity requirements more quickly by selecting segments with larger number of mobile users. Figure 4(c) displays the average graph density which falls as we increase the s values. This is due to larger cloaked subgraphs being formed for large s values. Again the mobility-aware approaches outperform the baseline randomized expansion and network expansion approach on this metric, even for higher s values. Figure 4(d) shows that the relative spatial resolution falls with increasing s-anonymity levels. This is expected due to larger cloaked subgraphs being formed for larger s-anonymity levels.

Varying Maximum Spatial Resolution: Figure 5 displays the experimental results with varying maximum spatial resolution σ_s. We choose the maximum spatial resolution values using a Zipfian distribution with parameter 0.6 from the following set of range values: $[0.5, 1.5]km$, $[1.5, 2.5]km$ and $[2.5, 3.5]km$. Also, we set the k-anonymity level between $[10, 50]$ values chosen with a Zipfian distribution with parameter 0.6, s-



(a) Relative k-Anonymity Level (b) Relative s-Anonymity Level (c) Average Graph Density (d) Relative Spatial Resolution

Fig. 5: User-defined Privacy and QoS Metrics with Varying Maximum Spatial Resolution



(a) Varying k-Anonymity (b) Varying s-Anonymity (c) Varying Spatial Resolution (d) Varying Randomness Factor

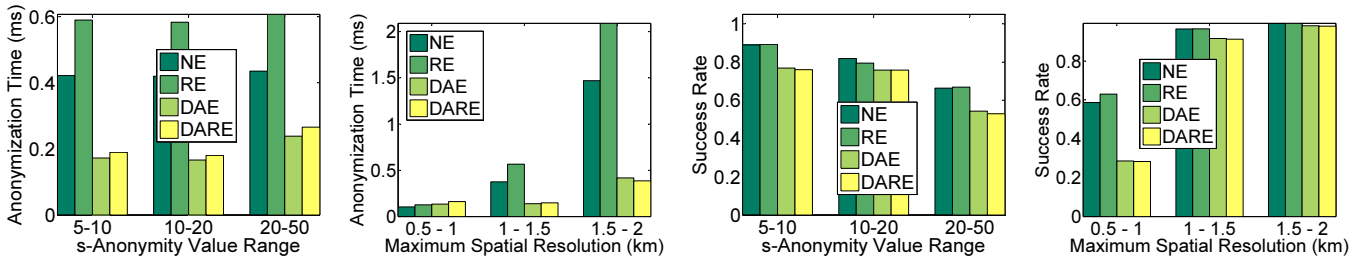
Fig. 6: Entropy Values for the Anonymization Algorithms

anonymity level between $[10, 20]$ values again chosen with a Zipfian distribution with parameter 0.6 and maximum temporal resolution with a mean value of 30 seconds.

Figure 5(a) shows that the relative k-anonymity levels are lower for lower maximum spatial resolution values. The mobility-aware approaches are constrained to select shorter segments with lower number of objects for low maximum spatial resolution. As the maximum spatial resolution restriction is relaxed, the relative k-anonymity levels increase, in particular for the mobility-aware approaches which select longer segments with larger number of mobile objects. On the other hand, the relative s-anonymity levels fall with increasing maximum spatial resolution as larger number of segments are selected due to relaxed spatial constraints (Figure 5(b)). The average graph density shows some interesting trends as observed in Figure 5(c). The randomized expansion approach experiences a drop in graph density with increasing maximum spatial resolution as more dispersed segments can be selected. The other approaches experience a slight increase in graph density with increasing maximum spatial resolution. This is due to the fact that as we increase the maximum spatial resolution, these approaches are able to select fewer segments with larger number of mobile users in order to construct a more compact cloaked subgraph. Figure 5(d) shows, as expected, that the relative spatial resolution increases with increasing maximum spatial resolution values.

2) Entropy-based Evaluation: In this experiment, we measure the entropy levels achieved by each approach. Figure 6(a) displays the entropy values with increasing k values; s values are chosen from the range $10 - 20$. As expected, the randomized expansion approach has the highest entropy levels, while the network expansion and the density-aware approach have relatively low entropy due to their deterministic nature. We see that introducing randomization to the density-aware approach results in much higher entropy levels; interestingly its entropy levels are competitive compared to the randomized approach. Figure 6(b) shows that entropy levels increase with increasing s values (k values in the range $10 - 50$) and density-aware randomized expansion has higher entropy values compared to the basic density-aware approach and network expansion. On the other hand, Figure 6(c) shows that as we increase the σ_s values, the entropy levels remain more or less stable. This shows that it is hard for adversaries to launch a successful attack on our density-aware randomized expansion approach, irrespective of the desired spatial resolution. Figure 6(d) displays the entropy values as we increase the value of the randomization factor α. As expected, entropy values increase sharply with increasing α. Furthermore, the segment entropy decreases as we increase the number of users in the system. This is due to fewer number of segments being selected to meet the k-anonymity requirements due to higher request density.

3) Performance Metrics: We observe two performance metrics for MOBICLOAK: anonymization success rate and average anonymization time with varying s-anonymity levels and varying σ_s. Figure 7(a) displays the average anonymization time measurements with increasing s values. The density-aware approaches are faster as they anonymize requests quickly by locating denser subgraphs. Similar trends are observed for increasing σ_s in Figure 7(b). The baseline randomized expansion approach requires more time to retrieve the segments in the pool and experiences high anonymization costs. The density-aware approaches have lower anonymization time, although it increases with increasing σ_s. Figure 7(c) displays that the success rate falls with increasing s values. This is due to the fact that meeting the higher s requirements with the same σ_s values becomes more difficult. Figure 7(d) shows that



(a) Anonymization Time with Varying s-Anonymity (b) Anonymization Time with Varying σ_s (c) Success Rate with Varying s-Anonymity (d) Success Rate with Varying σ_s

Fig. 7: Performance Metrics with Varying s-Anonymity and Spatial Resolution σ_s

the success rate increases with increasing σ_s values as larger number of requests can be anonymized. In addition, we also observe that the density-aware approaches have low success rate for very low spatial resolution values due to their tendency to select longer segments. However, for higher σ_s values their success rate is easily comparable to the naïve approaches.

V. Related Work and Conclusion

Existing anonymization solutions utilize location perturbation as a mechanism to disable an adversary from associating personally identifiable information with a location. Representative techniques for anonymization-based solutions are spatial cloaking [1], [6], [7], [8], [15], false dummies [12] and landmark objects [10], with spatial cloaking being the most popular protection mechanism. The criterion of transformation is solely based on location anonymity, and the amount of protection measured in terms of the area of the cloaked region.

There have been limited efforts on providing location privacy for users on road networks. XStar [20] is the first in-depth work that combines location k-anonymity with l-segment diversity, using star-based cloaking. However, as we pointed out in Section I, the l-segment diversity definition in [20] lacks graph density consideration and explicit support of the reciprocity criterion for successful cloaking. To the best of our knowledge, only CliqueCloak [6], NAP [16] and Prive [7] to date have made reciprocity as a mandatory criterion for successful cloaking.

Processing spatial queries over road networks has been an emerging research topic [19], [13], [3], [17]. The cost of query evaluation is a function of the set of segments and set of nodes considered for evaluation; thus, the MobiCloak approach is bound to have lowest query evaluation costs.

We have presented MobiCloak, a road network-aware location anonymization model for protecting location privacy under road network mobility models. We enrich user k-anonymity by introducing segment s-anonymity as a companion metric for guarding location privacy of mobile users on road networks. Additionally, we promote the use of graph density as an important measure for determining optimal cloaking subgraphs in a road network. Further, we devise a suite of graph-based cloaking algorithms which guarantee reciprocity - an important location cloaking property that many existing approaches fail to support - under an enhanced privacy model. Further, we use controlled randomization in the cloaking process to provide higher privacy strength while maintaining the utility of the cloaked location.

Acknowledgment

The last two authors were sponsored partially by the National Science Foundation under Grants IIS-0905493, CNS-1115375 and IIP-1230740.

References

[1] B. Bamba, L. Liu, P. Pesti, and T. Wang. Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. In *WWW*, 2008.
[2] A. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *Pervasive Computing, IEEE*, 2003.
[3] H. Cho and C. Chung. An Efficient and Scalable Approach to CNN Queries in a Road Network. In *VLDB*, 2005.
[4] M. Duckham and L. Kulik. A Formal Model of Obfuscation and Negotiation for Location Privacy. In *Pervasive*, pages 152–170, 2005.
[5] Freudiger, J. and Raya, M. and Félegyházi, M. and Papadimitratos, P. and Hubaux, J.P. Mix-Zones for Location Privacy in Vehicular Networks. In *Win-ITS*, 2007.
[6] B. Gedik and L. Liu. Location Privacy in Mobile Systems: A Personalized Anonymization Model. In *ICDCS*, 2005.
[7] G. Ghinita, P. Kalnis, and S. Skiadopoulos. PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems. In *WWW*, 2007.
[8] M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *MobiSys*, 2003.
[9] M. Gruteser and D. Grunwald. Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis. *Mobile Networks and Applications*, 2005.
[10] J. Hong and J. Landay. An Architecture for Privacy-Sensitive Ubiquitous Computing. In *MobiSys*, pages 177–189, 2004.
[11] P. Karger and Y. Frankel. Security and Privacy Threats to ITS. In *World Congress on Intelligent Transport Systems*, pages 2452–2458, 1995.
[12] H. Kido, Y. Yanagisawa, and T. Satoh. An Anonymous Communication Technique using Dummies for Location-based Services. In *IEEE ICPS*, pages 88–97, 2005.
[13] M. Kolahdouzan and C. Shahabi. Voronoi-based k Nearest Neighbor Search for Spatial Network Databases. In *VLDB*, 2004.
[14] A. Meyerson and R. Williams. On the Complexity of Optimal K-Anonymity. In *PODS*, 2004.
[15] M. Mokbel, C. Chow, and W. Aref. The New Casper: Query Processing for Location Services without Compromising Privacy. In *VLDB*, 2006.
[16] K. Mouratidis and M. Yiu. Anonymous Query Processing in Road Networks. *IEEE Transactions on Knowledge and Data Engineering*, pages 2–15, 2009.
[17] K. Mouratidis, M. Yiu, D. Papadias, and N. Mamoulis. Continuous Nearest Neighbor Monitoring in Road Networks. In *VLDB*, 2006.
[18] B. Palanisamy and L. Liu. Attack-Resilient Mix-Zones over Road Networks: Architecture and Algorithms. *IEEE Transactions on Mobile Computing*, pages 493–508, 2014.
[19] D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao. Query Processing in Spatial Network Databases. In *VLDB*, 2003.
[20] Wang, T. and Liu, L. Privacy-Aware Mobile Services over Road Networks. In *VLDB*, 2009.