Preserving Privacy in GPS Traces via Uncertainty-Aware Path Cloaking

Baik Hoh, Marco Gruteser
WINLAB / ECE Dept., Rutgers Univ.
Piscataway, NJ USA
baikhoh,gruteser@winlab.rutgers.edu

Hui Xiong
MSIS Dept., Rutgers Univ.
Newark, NJ USA
hui@rbs.rutgers.edu

Ansaf Alrabady
General Motors Corporation
Warren, MI USA
ansaf.alrabady@gm.com

ABSTRACT

Motivated by a probe-vehicle based automotive traffic monitoring system, this paper considers the problem of guaranteed anonymity in a dataset of location traces while maintaining high data accuracy. We find through analysis of a set of GPS traces from 233 vehicles that known privacy algorithms cannot meet accuracy requirements or fail to provide privacy guarantees for drivers in low-density areas. To overcome these challenges, we develop a novel time-to-confusion criterion to characterize privacy in a location dataset and propose an uncertainty-aware path cloaking algorithm that hides location samples in a dataset to provide a time-to-confusion guarantee for all vehicles. We show that this approach effectively guarantees worst case tracking bounds, while achieving significant data accuracy improvements.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy IssuesPrivacy; K.6 [**Management of Computing and Information Systems**]: Security and Protection

General Terms

Algorithms, Measurements, Security

Keywords

Privacy, GPS, Traffic

1. INTRODUCTION

A new class of applications that mines aggregate location traces from large numbers of users, spawned by the increasing ubiquity of sensors and wireless communications, raises significant privacy concerns. One example and the motivation for this paper is automotive traffic monitoring through probe vehicles [20, 14, 30], which infers traffic congestion from position and speed information periodically reported from GPS-equipped vehicles. Other applications of such aggregate location traces are road and city planning.

Privacy could be protected in such applications by rendering the data anonymous before sharing it with application service providers. An anonymous location dataset provides strong privacy protection while allowing sharing with arbitrary data consumers, since no purpose-binding restricts the data for certain uses. Anonymization, however, requires techniques beyond omitting obvious identifiers, since the spatio-temporal characteristics of the data allows tracking and reidentification of anonymous vehicles when user density is low. Existing algorithms based on the k-anonymity concept [42, 22, 24], however, modify the location traces substantially and cannot meet the accuracy requirements of the traffic monitoring application. Other techniques [6, 27, 36] achieve better accuracy but cannot guarantee privacy in low user density scenarios.

This paper addresses the challenge of providing strong privacy guarantees while maintaining high data accuracy of time-series location data. Specifically, the key contributions of this work are:

- introduction of a novel time-to-confusion metric to evaluate privacy in a set of location traces. This metric describes how long an individual vehicle can be tracked.
- development of an uncertainty-aware privacy algorithm that can guarantee a specified maximum time-to-confusion.
- demonstration through experiments on real-world GPS traces that this algorithm limits maximum time-to-confusion while providing more accurate location data than a random sampling baseline algorithm.

Overview. The remainder of this paper is organized as follows. Section 2 introduces the motivating traffic monitoring system and data requirements. In section 3, we describe the privacy model and introduce the time-to-confusion metric. Section 4 presents the uncertainty-aware privacy algorithm. Section 5 describes the experimental results obtained with real-world location traces, which demonstrate the privacy and data accuracy advantages. We then discuss limitations and extensions in section 6, review related work in section 7, and conclude.

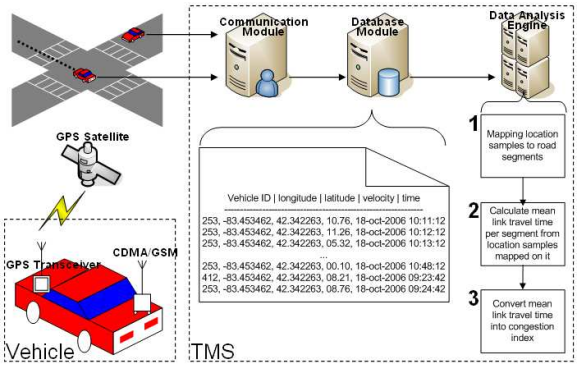
2. TRAFFIC MONITORING WITH PROBE VEHICLES

The traffic monitoring application that serves as a case study aims to provide estimates of current travel time for each route using real-time GPS traces from vehicles on these roads. The probe vehicles use on-board GPS receivers and cellular communications to periodically report their position and speed to a central traffic monitoring system, which stores them in a database for real-time and historical traffic analysis. Figure 1(a) illustrates this architecture.

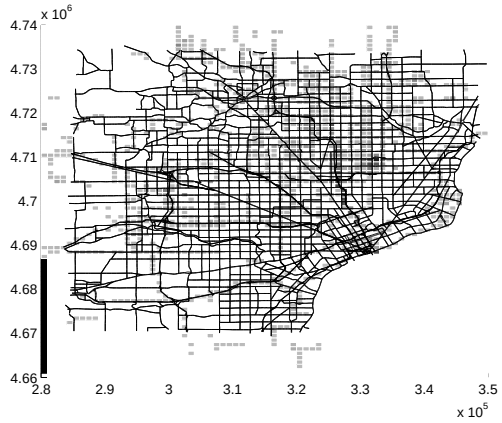
2.1 Real-world GPS Trace Collection

To obtain a realistic dataset similar to real deployments [2, 3, 1], we have offline collected a dataset containing GPS traces from

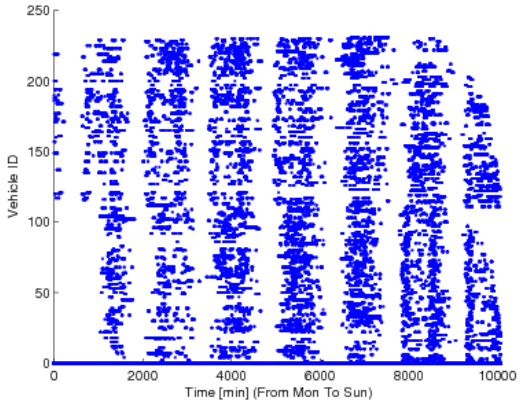
CCS'07, October 29–November 2, 2007, Alexandria, Virginia, USA.



(a) System architecture



(b) 70km x 70km road network with cell weights indicating the busiest areas



(c) Temporal distribution of GPS traces for 233 vehicles

Figure 1: Traffic monitoring system and spatio-temporal distribution of real-world dataset

233 vehicles driving in a large US city and its suburban area. Figure 1(b) depicts the 70 km by 70 km region that the vehicles covered. For privacy reasons no specific information about the vehicles or drivers is known to the authors, except that traces were recorded largely from test vehicles driven by volunteers. Each GPS sample comprises vehicle ID, timestamp, longitude, latitude, velocity, and heading information. Each vehicle records a GPS sample every minute, while its ignition is switched on, for a period of one week. This means that the traces contain temporal gaps, since no data is provided while the vehicle is parked with its ignition switched off. In addition, data is unavailable when the GPS reception is lost (e.g., due to obstruction from high-rise buildings) or the receiver is still in the process of acquiring the satellite fix. Figure 1(c) illustrates the distribution of gaps in the traces of around 233 vehicles. Each dot represents a received data sample. Since the traces do not contain information about ignition status, we assume that a gap longer than 10 min indicates that the vehicle was parked. We refer to the parts of a trace between two gaps longer than 10 min as a *trip*.

Parameter	Requirement
Spatial Accuracy	100m
Sample Interval	1min
Delay	few minutes

Table 1: Traffic monitoring system data requirements

2.2 Data Quality Metrics and Requirements

Data privacy algorithms increase privacy through deliberate modifications on the dataset, such as omission, perturbation, or generalization of a datum. Thus, there exists a tradeoff between data quality (or its utility) and the degree of privacy. To enable a meaningful evaluation of privacy algorithms let us discuss data quality requirements and metrics for the traffic monitoring application.

This application represents a road map as a graph comprising a set of road segments, where each road segment describes a stretch of road between two intersections. Generating the congestion map then proceeds in three steps: Mapping new GPS samples to road segments, computing mean road segment speed, and inferring a congestion index (e.g., by comparing current mean speed with nominal segment speed).

Mapping GPS samples onto road segments requires high *spatial accuracy*. Consider that two different parallel road segments (with traffic flow in same direction) may be only about 10m apart, as on the New Jersey Turnpike, for example. Cayford and Johnson [9] showed, however, that using tracking algorithms the correct road can be determined in 98.4% of all surface streets and 98.9% of all freeways if the location system provides a spatial accuracy of 100m and updates in 1s intervals. When reducing the update interval from 1s to 45s, the correctly determined roads drop from 99.5% to 98% (at 50m spatial accuracy). Therefore, to maintain high road mapping accuracy at the 1min sample interval for our data traces, we can assume that a minimum spatial accuracy of 100 m is needed.

Another important data quality requirement is *road coverage*, which primarily depends on the penetration rate, the percentage of vehicles carrying the traffic monitoring equipment. To achieve high coverage these systems aim at a minimum penetration rate of 3 (for freeways) to 5% (for surface streets) [14], but during the initial deployment phase penetration rates may be much lower. Thus privacy algorithms must offer protection even with in low deployment densities. Road coverage can also be reduced through privacy algorithms. Thus, we measure a *relative weighted coverage metric* for the privacy algorithms, which is based on the following heuristics. First, road coverage decreases as more samples are withheld. Second, probe-vehicle based traffic monitoring aims to extend traffic monitoring beyond a few key routes, but information from busier roadways is certainly more important than from low-traffic routes. Third, coverage is fundamentally limited by the number of probe vehicles on roads, thus we only consider coverage relative to the original dataset.

To measure the effect of removed samples on road coverage, relative weighted coverage first assigns each location-sample a weight,

depending on how busy the area around this sample is. Then, it divides the sum of weighted location samples from modified (or partially removed) traces by the sum of weighted location samples from the original traces. To estimate these weights for our dataset we divide the area into 1km by 1km grid cells and count the number of location samples n_i emanating from each cell i over one day in the original traces. The resulting weights for each cell are overlaid on the road map in Figure 1(b). The weights are normalized with the sum of weights over all samples, so that the relative weighted road coverage for the original dataset is equal to 1. More precisely, the weight for all samples in cell i equals $w_i = \frac{n_i}{\sum_j n_j^2}$. With these weights, relative weighted road coverage for a set of location samples L is then defined as $\sum_{l \in L} w_{c(l)}$, where the function c returns the cell index in which the specified location sample lies.

In summary, we can measure data quality for a traffic monitoring application through the relative weighted road coverage, where we consider a road segment covered if a data sample with sub-100m accuracy is available. Table 1 summarize key system parameters and requirements that we will assume in the following sections.

3. PRIVACY LEAKAGE THROUGH ANONYMOUS LOCATION TRACES

Especially in the United States where people rely heavily on automobiles and distances between buildings are large, monitoring the movements of a person's automobile can reveal sensitive information. First, knowing trip destinations can reveal information about a persons health, lifestyle, net worth, or political associations, Second, many drivers might object to such monitoring because it could reveal minor traffic or parking violations.

Even after anonymization, some of this information may be recovered, as simply removing identifiers from a dataset does not always provide strong anonymity guarantees, which was the motivation for introducing the k-anonymity concept [42].

3.1 Existing Privacy Algorithms

Several techniques have been proposed to increase location privacy. However, we are aware of only one class of techniques, spatial cloaking algorithms for k-anonymity, that can guarantee a defined degree of anonymity for all users.

k-anonymity [42, 38] formalizes the notion of strong anonymity and complementary algorithms exist to anonymize database tables. They key idea underlying these algorithms is to generalize a data record until it is indistinguishable from the records of at least $k-1$ other individuals. Specifically, for location information, spatial cloaking algorithms have been proposed [24, 22] that reduce the spatial accuracy of each location sample until it meets the k-anonymity contstraint. To achieve this, the algorithms require knowledge of nearby vehicles positions, thus they are usually implemented on a trusted server with access to all vehicles current position.

k-anonymous datasets produced with known algorithms cannot meet traffic monitoring's accuracy requirements. Figure 2 shows the spatial accuracy results obtained after applying a spatial cloaking algorithm to guarantee k-anonymity of each sample. We use the same dataset in section 5.1 so that we could directly compare k-anonymity with our proposed solution in terms of spatial accuracy. The results were obtained with the CliqueCloak algorithm [22], which to our knowledge achieves the best accuracy. *The results show that even for very low privacy settings, $k = 3$, location error remains close to 1000m for an emulated deployment of 2000 vehicles, far over the accuracy requirement of the traffic monitoring application.* While these results can be expected to improve with increased penetration rates as the deployment case of 5500 vehicles shows 500m for $k = 3$ (indeed, [24] shows that median accuracies of 125 meters and below can be obtained when *all* vehicles act as probes), other privacy approaches are necessary to enable probe systems operating with lower penetration rates.

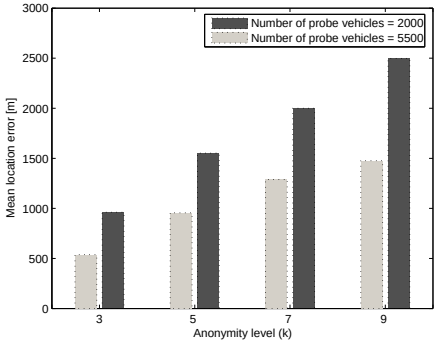


Figure 2: Data accuracy of samples processed with spatial cloaking algorithm fails to meet the accuracy requirement in our scenario

Best effort algorithms. Given that in dense environments paths from many drivers cross, drivers intuitively enjoy a degree of anonymity, similar to that of a person walking through an inner-city crowd. Thus, Tang et al. [43] lay out a set of privacy guidelines and suggest that the sampling frequency, with which probes send position updates, should be limited to larger intervals. The authors mention that a sample interval of 10min appears suitable to maintain privacy, although the choice appears somewhat arbitrary (for reference, a typical consumer GPS chipset implementation offers a maximum sampling frequency of 1 Hz). We refer to data collection with reduced sampling frequency as subsampling.

Other best effort algorithms suppress information only in certain high-density areas rather than uniformly over the traces as the subsampling approach. The motivation for these algorithms that path suppression in high density areas increases the chance for confusing or mixing several different traces. This approach was first proposed by Beresford and Stajano [7]. The path confusion [27] algorithm also concentrates on such high-density areas although it perturbs location samples rather than suppressing them. These techniques increase the chance of confusion in high-density areas, but they also cannot guarantee strong privacy in low-density areas where paths only infrequently meet. Thus, in-terms of worst-case privacy guarantees their advantage over subsampling remains unclear.

We choose the subsampling algorithm as a best effort baseline algorithm. Table 2 shows an adversary's tracking performance over an anonymous set of samples with 1 min (no removal) and 2 min (50% removal) sampling intervals. For a probe vehicle density of 500 vehicles per a 70km^2 region, the tracking algorithm returns 3480 segments of 15 min duration and 1172 segments of 20 min duration. Both reducing the sampling interval and increasing probe vehicle density reduces tracking performance. For example, with 2000 vehicles on a same area and 2 min sampling interval, 17 segments of 20 min duration can be identified. Precision of the tracking algorithm is about 95% in all cases, meaning that only 5% of the returned segments do not match an actual vehicles path, except in the 2000 vehicle 2 min case, where relatively few segments can be tracked (in this case precision drops to 60 percent). These example results were obtained with a tracking model that we will describe in detail in the following section.

	Random sampling (50% removal)		Anonymization (no removal)	
	15min	20min	15min	20min
Density=500, Uncertainty threshold=0.45	45/47	28/29	3300/3480	1117/1172
Density=2000, Uncertainty threshold=0.7	18/30	10/17	1302/1394	908/958

Table 2: Empirical confidence in subsampling

To understand the implications of these tracking durations (15min and 20 min), let us consider figure 3, which depicts the histogram of per-trip travel time in the GPS dataset. The data shows a large number of very short trips, for example 30% of trips are shorter than 10 min, 50% of trips shorter than 18min. This empirical result also coincides with the empirical statistics from real GPS traces in Krumm's work [35] (Krumm observes 14.4 min per trip as a median). This means that by following a trace for only 10min, an adversary may be able to track a vehicle from its home to a sensitive destination.

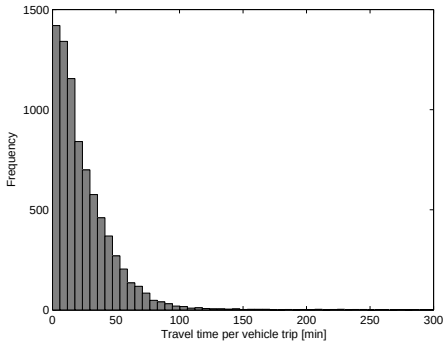


Figure 3: Empirical distribution of travel times per vehicle trip.

These results illustrate that protecting *all* drivers of probe vehicles through subsampling remains difficult. One minute sampling intervals are already large for a traffic monitoring application but protecting all drivers even in low density areas would require a further significant increase in the sampling interval. Moreover, it is difficult to choose this sampling interval since traffic densities can change substantially over time and space.

This raises the question of alternate definitions and measures for anonymity in location traces.

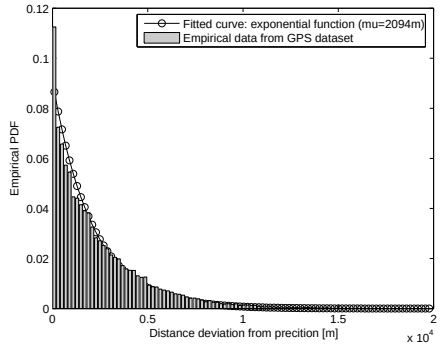


Figure 4: Fitting distance errors in tracking using an exponential function

3.2 Privacy Metric and Adversary Model

We observe that the degree of privacy risk strongly depends on how long an adversary can follow a vehicle. To constitute a privacy breach a trace must contain a privacy sensitive event (e.g., visited a sensitive destination) and the adversary must be able to identify the driver generating this trace. Both the probability that sensitive information is included and the probability of identification increase with longer traces. Identification may be possible, for example, if the vehicle returns to a known home or work location of a specific individual.

Since consecutive location samples from a vehicle exhibit temporal and spatial correlation, paths of individual vehicles can be reconstructed from a mix of anonymous samples belonging to several vehicles. This process can be formalized and automated through target tracking algorithms [26]. These algorithms generally predict the target position using the last known speed and heading information and then decide which next sample to link to the same vehicle through Maximum Likelihood Detection [44]. If multiple candidate samples exist, the algorithm chooses the one with the highest a posteriori probability based on a probability model of distance and time deviations from the prediction (in our evaluation, we assume a strong adversary with a good model of these deviations). If several of these samples appear similarly likely, no decision with high certainty is possible and tracking stops.

Privacy Metrics. In consideration of this adversary model, we measure the degree of privacy as the *Mean Time To Confusion (MTTC)*, the time that an adversary could correctly follow a trace. Note that this includes time while a user remains stationary unless otherwise specified. More specifically, the time to confusion is the tracking time between two points where the adversary reached confusion (i.e., could not determine the next sample with sufficient certainty). Inspired by the use of entropy in anonymous communication systems [40, 16], we use information theoretic metrics to measure uncertainty or confusion in tracking.

For any point on the trace, *Tracking Uncertainty* is defined as $H = -\sum p_i \log p_i$, where p_i denotes the probability that location sample i belongs to the vehicle currently tracked. Lower values of H indicate more certainty or lower privacy. Given no other information than the set of location samples, intuitively the probability for a sample reported at time t is high, if the sample lies close to the predicted position of the vehicle at time t and if no other samples at the same time are close to the vehicle. As one step further, we can also express tracking confidence C on adversary's trial by calculating $(1 - H)$.

Empirically, we found that distances of the correct sample to the predicted position appear monotonically decreasing in figure 4. Therefore, we compute the probability p_i for a given location sample by first evaluating the exponential function

$$\hat{p_i} = e^{-\frac{d_i}{\mu}}$$

for every candidate sample and then normalizing all $\hat{p_i}$ to obtain p_i. The parameter μ can be interpreted as a distance difference that can be considered very significant. We obtain the value of μ from empirical pdf of distance deviation in figure 4 which we fit with

exponential function using unconstrained nonlinear minimization (μ is 2094 meters).

The following algorithm is not dependent on the use of an exponential function for estimating the probability that a location sample belongs to the same trace. It does assume, however, that a publicly-known 'best' tracking model exists and that the adversary does not have any better tracking capabilities. In this paper, we have empirically derived this probability model by fitting an exponential function.

Overall, the mean time to confusion can then be defined as the mean tracking time during which uncertainty stays below a confusion threshold. If the uncertainty threshold is chosen high, tracking times increase but so also does the number of false positives (following incorrect traces). Since the adversary cannot easily distinguish correct tracks and false positives, we assume that high uncertainty thresholds will be used.

4. PATH PRIVACY-PRESERVING MECHANISM

In this section, we present a method for preserving privacy in GPS traces that can guarantee a level of privacy even for users driving in low-density areas. Given a maximum allowable time to confusion and an associated uncertainty threshold, the algorithm can process a stream of received position samples to maintain the tracking time bounds.

Since the algorithm must be aware of the positions of other vehicles, we first develop a centralized solution and then discuss how reliance on a trustworthy privacy server may be relaxed. We first consider the stepwise tracking model without the possibility of path reacquisition.

We observe that a specified maximum time to confusion (for a given uncertainty level) can be guaranteed if the algorithm only reveals location samples when (i) time since the last point of confusion is less than the maximum specified time to confusion or (ii) at the current time tracking uncertainty is above the threshold.

Algorithm 1 shows how this idea can be implemented. Note that it describes processing of data from a single time interval, it would be repeated for each subsequent time slot with the state in the vehicle objects maintained. It takes as input the set of GPS samples reported at time t (v.currentGPSSample updated for each vehicle), the maximum time to confusion (confusionTimeout), and the associated uncertainty threshold (confusionLevel). Its output is a set of GPS samples that can be published while maintaining the specified privacy guarantees.

The algorithm proceeds as follows. It first identifies the vehicles that can be safely revealed because less time than confusionTimeout has passed since the last point of confusion (line 12f.) Second, it identifies a set of vehicles that can be revealed because current tracking uncertainty is higher than specified in confusionLevel (line 15-30). Finally, it updates the time of the last confusion point and the last visible GPS sample for each vehicle (line 32ff., the latter is needed for path prediction in the uncertainty calculation). This step can only be performed when the set of revealed GPS samples had been decided, since confusion should only be calculated over the revealed samples.

The second step relies on several approximations. To reduce computational complexity it calculates tracking uncertainty only with the k closest samples to the prediction point, rather than with all samples reported at time t. This is a conservative approximation, since uncertainty would increase if additional samples are taken into account (see proof in appendix A). Further, it builds a set of releaseCandidates since uncertainty should only be calculated with released samples, but the set of released samples is not determined yet. The algorithm subsequently prunes the candidate set until only vehicles remain who meet the uncertainty threshold. The key property to achieve after the pruning step is that $\forall$ v $\in$ releaseCandidates. uncertainty(v.predictedPos, k closest neighbors in releaseSet $\cup$ releaseCandidates) $\geq$ confusionLevel. The algorithm uses the approximation of calculating the k closest neighbors before the pruning phase, and ensuring during pruning that only vehicles remain if all k neighbors are in the set. While this approximation could be improved in order to release more samples, the current version is sufficient to maintain the privacy guarantee.

Algorithm 1 Uncertainty-aware privacy algorithm

```
 1: // Determines which location samples can be release while maintaining
    privacy guarantee.
 2: releaseSet = releaseCandidates = {}
 3: for all vehicles v do
 4:    if start of trip then
 5:       v.lastConfusionTime = t
 6:    else
 7:       v.predictedPos = v.lastVisible.position +
 8:       (t-v.lastVisible.time)*v.LastVisible.speed
 9:    end if
10:
11:    // release all vehicles below time to confusion threshold
12:    if t - v.LastConfusionTime < confusionTimeout then
13:       add v to releaseSet
14:    else
15:       // consider release of others dependent on uncertainty
16:       v.dependencies = k vehicles closest to the predictedPos
17:       if uncertainty(v.predictedPos, v.dependencies) > confusionLevel
          then
18:          add v to releaseCandidates
19:       end if
20:    end if
21: end for
22:
23: // prune releaseCandidates
24: for all  v ∈ releaseCandidates do
25:    if ∃ w ∈ v.dependencies. w ∋ releaseCandidates ∪ releaseSet then
26:       delete v from releaseCandidates
27:    end if
28: end for
29: repeat pruning until no more candidates to remove
30: releaseSet = releaseSet ∪ releaseCandidates
31:
32: // release GPS samples and update time of confusion
33: for all v ∈ releaseSet do
34:    publish v.currentGPSSample
35:    v.lastVisible = v.currentGPSSample
36:    neighbors = k closest vehicles to v.predictedPos in releaseSet
37:    if uncertainty(v.predictedPos, neighbors) >= confusionLevel then
38:       v.lastConfusionTime=t
39:    end if
40: end for
```

4.1 Algorithm Extensions for the Reacquisition Tracking Model

The algorithm described so far does not provide adequate privacy guarantees under the reacquisition tracking model because it only ensures a single point of confusion after the maximum time to confusion has expired. Recall that under the reacquisition model an adversary skips samples with high confusion under certain conditions and thus may be able to reacquire the correct trace even after a point of confusion.

We observe that such reacquisitions are only possible over short time-scales, since movements after more than several minutes become too unpredictable. To verify this assumption, figure 5 shows

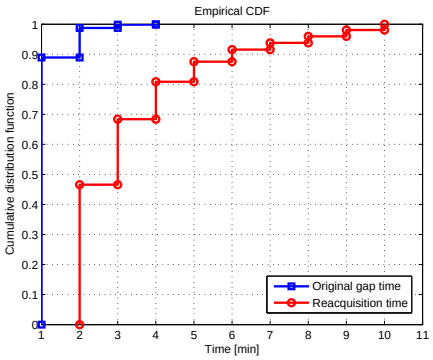


Figure 5: Cumulative distribution function of reacquisitions

the longest reacquisition and distribution of reacquisition length in minutes, empirically obtained from our dataset. As expected, no reacquisitions occur over gaps longer than 10 minutes. Thus, the following extensions can prevent reacquisitions within a time window w. For the experiments reported in the following section we set $w = 10$.

- **After the *confusionTimeout* expires:** In addition to maintaining confusion from the last released position, it is calculated from every prior released location sample (of the same vehicle) within the last w minutes. Samples can only be released if all these confusion values are above the confusion threshold.
- **Before the *confusionTimeout* expires:** Every released sample must maintain confusion to any samples which are released during the last w minutes *and* before the *confusionTimeout* was last reset.

5. EXPERIMENTAL EVALUATION

In this section, we present the experimental evaluation of the proposed privacy preserving techniques. Specifically, we demonstrate: (1) the effectiveness of our proposed techniques for privacy protection in the analysis of GPS traces; (2) how our proposed privacy preserving techniques can maintain the quality-of-service for the traffic monitoring application.

5.1 Experimental Setup

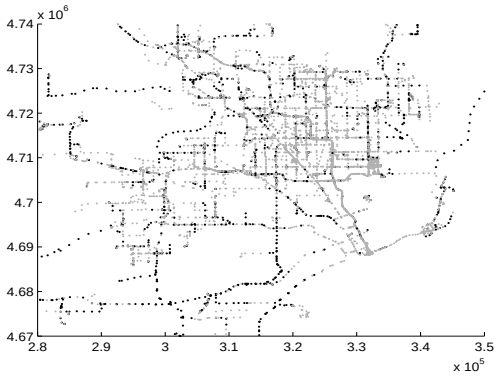
Experimental Data Sets. In our experiments, we used trace-driven simulations for capturing real vehicle movements, density, GPS inaccuracies, and road network artifacts. In the experiments, we first applied privacy preserving techniques on the GPS traces and then tested the performance of privacy protection using target tracking techniques on these privacy-preserved GPS traces.

Since target tracking typically is only effective for a short time period, we conducted the targeting tracking experiments on 24-hour GPS traces in two different user density scenarios: 500 probe vehicles and 2000 probe vehicles on a 70km^2 region. To create a high density scenario, we overlay GPS Traces of different volunteer drivers at the same time frame (24 hours) of different dates. A limitation of this overlay method is that it generates similar routes by aggregating GPS traces from the same set of drivers. Still, we believe that it provides insights into higher density deployments (we will revisit this limitation in the discussion section.)

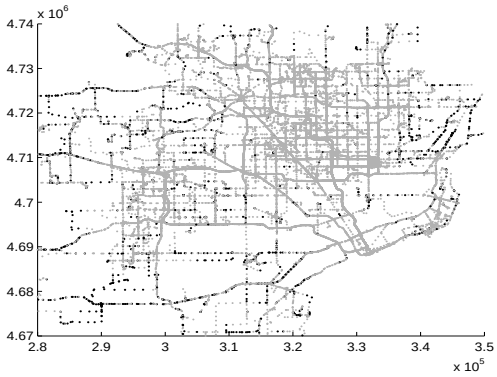
Evaluation Metrics. In our experiments, we applied the following two metrics to evaluate our privacy preserving algorithms for GPS traces.

Tracking Time. Minimizing tracking time reduces the risk that an adversary can correlate an identity with sensitive locations. We use *time to confusion (TTC)*, which we defined in section 3 as a privacy metric, to measure the tracking duration. To better demonstrate the bounded privacy protection of our proposed algorithm, we report two statistics: the maximum value of TTC and the median value of TTC.

(Relative) Weighted Road Coverage. This metric provides an indication of data quality for the traffic monitoring applications. Indeed, there is a tradeoff between privacy protection and the flexible use of data. In our case, a privacy preserving algorithm must provide reasonable privacy protection while delivering the same road coverage for satisfying the need of the traffic monitoring applications. In this paper, we use *relative* road coverage as we defined in section 2. In addition to this metric, we also provide the percentage of released location sample compared to the original traces which we consider 100%. Note that both metrics are normalized by values of the original GPS traces.



(a) Snapshot of privacy-preserving GPS traces generated by uncertainty-aware path cloaking at off-peak time (over 1.5 hours) in a high density scenario



(b) Snapshot of privacy-preserving GPS traces generated by uncertainty-aware path cloaking algorithm at peak time (over 1.5 hour) in a high density scenario

Figure 6: Uncertainty-aware privacy algorithm removes more samples in low-density areas, in which vehicles could be easily tracked. Gray dots indicate released location samples, black ones denote removed samples.

Snapshots of Privacy-preserving GPS Traces. Let us compare the privacy-preserving GPS traces generated by the proposed path cloaking algorithm with the original GPS traces. Figures 6(a) and 6(b) show both in a high user density scenario for off-peak

(over 1.5 hours at 10am) and peak time (over 1.5 hour at 5pm), respectively. Gray dots indicate released location samples while black dots illustrate samples removed by path cloaking. We observe two characteristics from these traces. First, uncertainty-aware path cloaking removes fewer location samples at peak time and second, it retains more location samples within the presumably busier downtown area. This illustrates how the algorithm, by virtue of its design, retains information on busier roads where traffic information is most valuable.

5.2 Protection Against Target Tracking

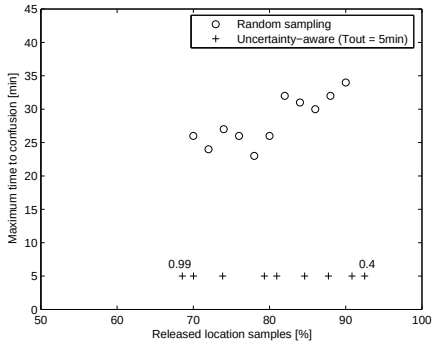
The following target tracking experiment illustrates how the path cloaking algorithm prevents an adversary from reconstructing an individual's path using the cleansed GPS traces. Specifically, we compare our uncertainty-aware privacy algorithm and its *with-reacquisition* version with random subsampling in terms of maximum and median TTC for configurations that produce the same number of released location samples (as a metric of data quality). We evaluate the effectiveness of our proposed privacy preserving algorithms by answering the following questions:

- Do uncertainty-aware privacy algorithms effectively limit tracking time (i.e., guarantee time-to-confusion)? Are these limits maintained even in low-user density scenarios?
- How does the average tracking time allowed by path cloaking compare to the subsampling baseline, at the same data quality level.
- How are the results affected by the choice of data quality metric (percentage of released location samples vs relative weighted road coverage)?

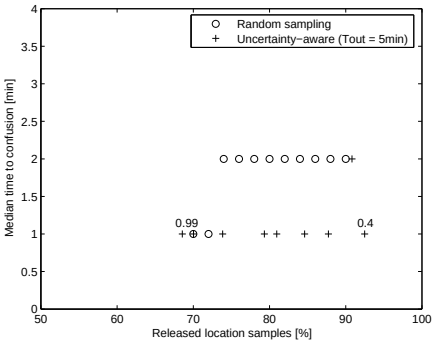
Throughout the results presented in the following subsections, one graph depicts many experiment trials, where one trial comprises the following steps. We first apply a privacy algorithm to the low-density (500 vehicle) or high-density (2000 vehicle) dataset generated from the 233 original vehicle traces. We then remove vehicle identifiers and execute the target tracking algorithm (see Sec. 3) to measure tracking time for the first 233 vehicles. For each vehicle, we compute the tracking time starting from each sample of the trace and report the maximum. One data point shown in the graph then corresponds to the median or maximum over the 233 vehicle tracking times computed for one trial. For each graph, these trials are then repeated with different uncertainty thresholds for the path cloaking algorithms and different probabilities of removal in the subsampling algorithm.

5.2.1 Bounded Tracking Time without Reacquisition

First, we ascertain whether the uncertainty-aware privacy algorithm guarantees bounded tracking under the no reacquisition tracking assumption . Figures 7(a) and 7(b) show the maximum and median tracking time plotted against the relative amount of released location samples, respectively, for a high density scenario with 2000 vehicles in the 70km-by-70km area. Figure 7(a) shows results for the uncertainty-aware privacy algorithm (marked with +) for varying uncertainty levels with timeout fixed at 5 minutes and for the random subsampling algorithm for varying probabilities of removal. Since the configuration parameters from these algorithms are not directly comparable, the graph shows the percentage of released location samples on the x-axis, allowing comparison of TTC at the same data quality level. Also note that graph compares the algorithms in terms of maximum tracking time, to illustrate differences in tracking time variance and outliers. During tracking we set the adversary's uncertainty threshold to 0.4. This means that the adversary will give up tracking if at any point the uncertainty level rises above this threshold, because the correct trace cannot be determined. A 0.4 uncertainty level corresponds to a minimum probability of 0.92 for the most probable next location sample.



(a) The Maximum Value of TTC using Uncertainty-aware privacy algorithm without Reacquisition



(b) The Median Value of TTC using Uncertainty-aware privacy algorithm without Reacquisition

Figure 7: Maximum / Median tracking duration for different privacy algorithms in high density scenarios (2000 vehicles / 1600 sqm). The Uncertainty-aware privacy algorithm outperforms random sampling for a given number of released location samples.

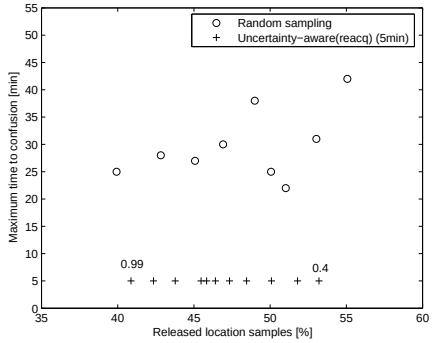
As evident from the data, the uncertainty-aware privacy algorithm effectively limits time to confusion to 5 min, except for very low privacy settings (i.e., low uncertainty threshold less than 0.4), while the random sampling algorithm allows some vehicles to be tracked up to about 35min. Our proposed algorithm can release up to 92.5% of original location samples while achieving the bounded tracking property.

In figure 7(b), we see that naturally occurring crossings and merges in the paths of nearby vehicles lowers median TTC to 1 or 2 minutes (with reacquisition it would be higher, though). However, with random subsampling (20% removal), about 15% of vehicles (34 out of 233) can still be tracked longer than 10 minutes. The uncertainty-aware path cloaking can guarantee the specified maximum tracking time of 5min even for these vehicles with higher data quality, removing only 17.5% of samples.

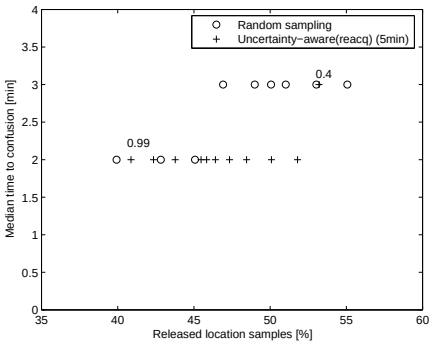
5.2.2 Dependence on Reacquisition and Density

We now repeat the same experiment under the reacquisition tracking model, where an adversary may skip ahead over a point of confusion. Figure 8(a) (note scaled x-axis) shows that the uncertainty-

aware privacy algorithm with reacquisition extensions can also effectively limit tracking time under this model, while subsampling allows a worst case tracking time of 42 min. Figure 8(b) also shows that the median tracking time is increased by one minute due to the change in tracking model. The maximum allowable amount of released location samples is decreased compared to that of figure 7.



(a) The Maximum Value of TTC using the (with reacquisition) Uncertainty-aware privacy algorithm



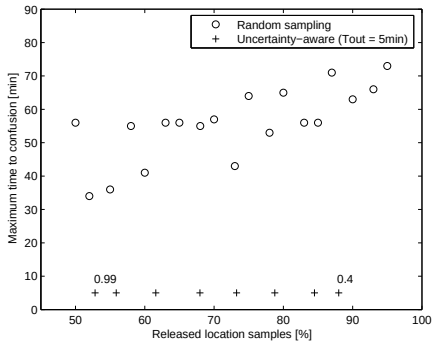
(b) The Median Value of TTC using the (with reacquisition) Uncertainty-aware privacy algorithm

Figure 8: Maximum / Median tracking duration for different privacy algorithms in high density scenarios (2000 vehicles / 1600 sqm) under the reacquisition tracking model.

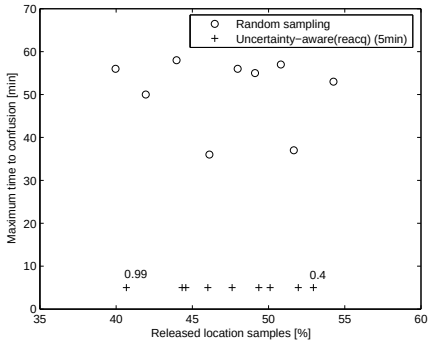
Let us now investigate whether the privacy guarantee is also maintained in a very low user density scenario with only 500 probe vehicles. Figure 9 shows that this is indeed the case both with and without the reacquisition model. While subsampling allows a longer maximum TTC due to the low user density, our proposed scheme still preserves the maximum TTC guarantee of 5 minutes by removing 1.8% to 14.8% more samples (for uncertainty thresholds between 0.4 and 0.99). The same result can be observed in figure 9(b) with reacquisition, except that the difference in samples removed is not as pronounced. Compared to the high density scenario, our proposed algorithm requires removing more samples to achieve the bounded tracking property in the lower user density scenario.

5.3 Quality of Service Analysis

So far, we have measured quality of service in terms of the percentage of samples removed by the algorithm. Since samples in higher density areas are more important for the traffic monitoring application, the benefits of our proposed privacy algorithm are even more significant if we consider *relative weighted road coverage*. More details are shown in figure 10. Figure 11(b) further shows that the uncertainty-aware privacy algorithm achieves a relative weighted road coverage similar to that of original location traces even though the actual number of released location samples is lower than that of original location traces as shown in figure 11(a). Figure 6 explains this results, in that the algorithm retains most samples in high-density areas and removes most from lower densities. However, the uncertainty-aware privacy algorithm with reacquisition extensions provides a slight improvement of relative QoS for weighted road coverage. More detailed statistics on this improvement are provided in table 3.



(a) The Maximum Value of TTC using the Uncertainty-aware privacy algorithm without Reacquisition



(b) The Maximum Value of TTC using the (with reacquisition) Uncertainty-aware privacy algorithm

Figure 9: The Uncertainty-aware privacy algorithm and its (with reacquisition) version outperform a random subsampling at a given range of sample removal also in the low density scenarios (500 vehicles / 1600 sqm).

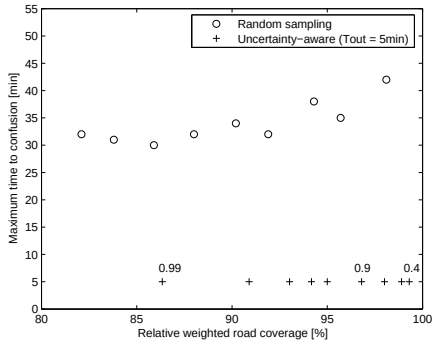
6. DISCUSSION

The following issues warrant a more detailed discussion.

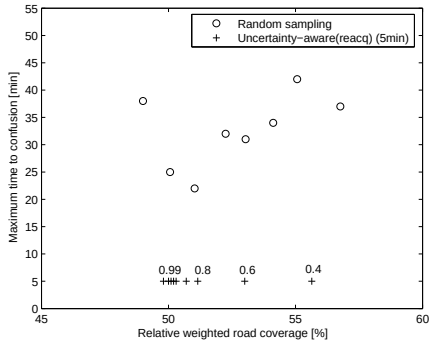
Map-based Tracking. Tracking performance would likely be improved by also considering road maps during the linking process. For example, the adversary could assign a lower probability to a segment if no direct road connection exists, even though the segment is near the predicted position. The algorithm could also adjust the predicted location based on actual roadway positions [11]. To counter this more sophisticated tracking, the bounded privacy algorithm could also take these road maps into account when computing entropy values. The complete analysis remains an open problem for future work.

	QoS metrics	
	Released location samples	Weighted road coverage
Original traces	100%	100%
Uncertainty-aware privacy (5min,0.95)	81%	95.0%
Random sampling (0.8)	80%	79.3%
(with reacq) Uncertainty-aware (5min,0.4)	53.2%	55.6%
Random sampling (0.53)	53%	52.9%

Table 3: Quality of service enhancement in each of Uncertainty-aware privacy algorithm, (with reacquisition) Uncertainty-aware privacy algorithm, and random sampling compared to the QoS level which original traces can achieve.

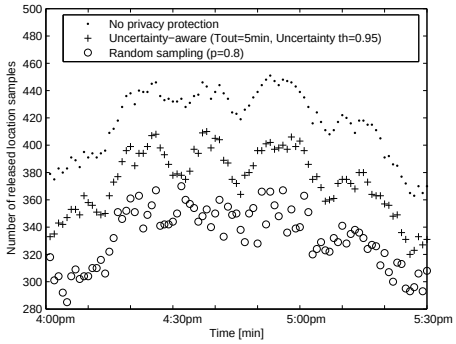


(a) Comparison of Maximum TTC against Weighted Road Coverage in High Density Scenario (Uncertainty-aware privacy algorithm)

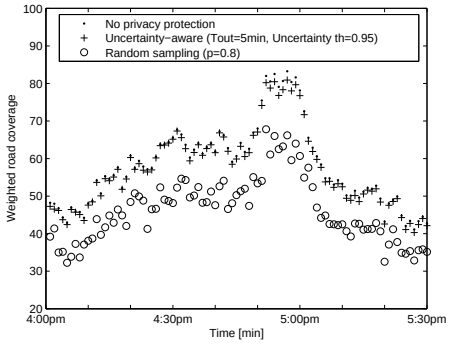


(b) Comparison of Maximum TTC against Weighted Road Coverage in High Density Scenario ((with reacquisition) Uncertainty-aware privacy algorithm)

Figure 10: Time-to-confusion advantages of uncertainty-aware path cloaking become even more pronounced when comparing algorithms with the traffic-monitoring-specific (Relative) Weighted Road Coverage data quality metric.



(a) Number of Released Location Samples in Peak Time



(b) Weighted Road Coverage in Peak Time

Figure 11: The Uncertainty-aware privacy algorithm removes more samples in low density area, leading to enhanced QoS in the high density regions, where traffic monitoring information is most valuable.

A priori knowledge. In this work, we have concentrated on data mining and inference techniques that do not possess any a priori knowledge about the individual drivers. Even if home identification and tracking in general remain difficult, an analyst could infer sensitive information by focusing on a select individual. For example, given the home and work position of an individual, it is possible to determine when the person left home and passed an accident site because the tracking analysis a priori knows the destination of the trip. The detailed analysis of this case also remains for future work.

Relaxing trust in location server. The described centralized algorithm requires a trustworthy location server, since the algorithm needs the full GPS traces of all vehicles. A fully distributed algorithm poses a research challenge by itself, since clients would need to monitor the positions of neighboring cars, which again raises privacy concerns. It also appears possible, though, to relax the trust assumptions in the location server through a hybrid approach, with additional in-vehicle disclosure control based on coarser information about neighbors. Since data quality would only be marginally affected by missing updates in low-density areas, one could devise schemes to inform vehicle of the approximate probe density in their area. Then vehicles could reduce location updates to the server in the most sensitive low-density areas. To prevent spoofing of such density information, further research could investigate data cross-validation schemes or secure multi-party computation schemes to compute density.

Dataset limitations. Finally, we need to point out that the tracking results can be affected by the choice of probe vehicles. In our dataset, most drivers shared the same workplace. Thus, the workplace acted as a place of confusion, where the tracking algorithms failed. A random sample of the population would probably not share such a common location, thus we would expect tracking performance to improve. This would cause both our proposed algorithms and the random sampling method to remove more samples to meet the maximum TTC. The performance gap between them might also change from what we have observed in our study. In addition, our method of overlaying multiple datasets to create one high-density scenario may not be entirely faithful in representing true traffic conditions. Due to this overlay, some of the vehicles may also be driven by the same driver on similar routes, creating a further bias towards reduced tracking performance. Nonetheless, we believe our current results provide a valuable first step towards understanding tracking performance in probe vehicle scenarios.

7. RELATED WORK

The question of data anonymity has been studied for some time (e.g., [38]) but a solution that achieves both strong privacy guarantees and a high degree of data accuracy for time-series location data remains elusive. Not surprisingly, recent analyses of GPS traces [34, 28, 26] have shown that simply omitting obvious identifiers from a dataset does not guarantee anonymity. Thus, stronger protection mechanisms are needed. Specifically, the k-anonymity concept [38, 42] has been adapted for location-based services [24, 37, 22]. If user density is high, these solutions can provide sufficient accuracy for applications such as point-of-interest queries, but as we have shown they do not achieve the high accuracy requirements of traffic monitoring applications with low penetration rates.

Similarly, random perturbation approaches for privacy-aware data mining [5, 4], which seek to modify a dataset to guarantee privacy of data subjects while preserving utility of the data, are not applicable in this context. Noise with large variance does not preserve sufficient data accuracy, while noise with small variance may be filtered by tracking algorithms due to the spatio-temporal nature of the data (in addition to the general weaknesses pointed out by Kargupta et al. [33]).

Thus, several best effort location data protection algorithms have been suggested [7, 39, 36, 27, 8], which have in common that they create areas of confusion where the traces from several users converge. While these algorithms achieve better accuracy and provide a defined level of privacy in such an area of confusion, they cannot provide overall privacy guarantees because these areas of confusion might not occur in lower-density areas.

Anonymity has also been extensively studied in the networking domain. Starting from Chaum's anonymous communication [10] work, researchers have developed MIX networks such as Onion Routing [23] or Tor [17]. Privacy of *location information* has been extensively investigated at the network-level. Network-level privacy techniques such as mixes and pseudonyms have been developed for cellular networks [19] and mobile IP [18]. The usage of silent periods [25, 39, 36, 29], periods of no communication, was proposed for wireless networks to reduce exposure to tracking. Sharing a similar approach with *swing & swap* by Li et al. [36], Jiang et al. combined three known concepts (silent period, pseudonym update, and control of transmission) to maximize the size of anonymity set in their work [31]. For sensor networks, two research groups, Kamat et al. and Deng et al. [32, 15] develop routing algorithms to protect the location of message senders or receivers (i.e., base station). These approaches are largely complementary to our work, they could be used in relaying (encrypted) GPS readings to the traffic monitoring service provider. The work on measuring communication anonymity [40, 16] also inspired us to use entropy in defining time to confusion.

Another proposed approach builds on privacy policy languages [13] and their location-oriented extensions [41] to allow users (or their automated agents) to make more informed decisions about data sharing. Such policies may be enforced through access control mechanisms, such as [21, 45] for spatio-temporal data. Using these approaches, data can only be shared if the data provider trusts the data consumer.

8. CONCLUSIONS

In this paper, we have proposed a novel time-to-confusion metric to characterize the degree of privacy in an anonymous set of location traces. We then developed an uncertainty-aware privacy algorithm, which can guarantee a defined maximum time-to-confusion for all vehicles, even those driving in low density areas. We showed through experiments with real-world GPS traces that the algorithm can effectively guarantee a maximum time-to-confusion, while a random sampling baseline algorithm allows tracking time outliers for vehicles in low density regions at the same data accuracy level.

APPENDIX

A. PROOF OF THEOREM

Theorem A. *Given n non-zero probabilities $p_0, p_1, \ldots p_n$, let $H(S_i)$ be the entropy calculated over the normalized probabilities of the $i \leq n$ most probable hypotheses. Then, $H(S_i) \leq H(S_n)$.*

PROOF. Let us order the probabilities so that $p_1 \geq p_2 \geq p_3 \geq ... \geq p_n$. We then refer to the set which includes the normalized probabilities from the first to the ith one $\frac{p_1}{\sum_i p_i}, \ldots \frac{p_i}{\sum_i p_i}$ as S_i. The entropy of S_1 is 0, since the event is certain, and thus $S_1 \leq S_2$. More generally, we know from [12] that the following relation holds between $H(S_i)$ and $H(S_{i+1})$.

$$\alpha H(p_1, p_2, ..., p_i) + H(\alpha, 1-\alpha) = H(\alpha p_1, \alpha p_2, ..., \alpha p_i, 1-\alpha) \quad (1)$$

Since we ordered the probabilities (descending) and $(1-\alpha)$ is the $(i+1)$th probability in S_{i+1}, we also know that $(1-\alpha) \leq \frac{1}{i+1}$. Thus, $\frac{i}{i+1} \leq \alpha \leq 1$ holds, given that $\alpha \leq 1$ as a probability. In terms of $H(S_i)$ and $H(S_{i+1})$ equation 1 can be rewritten as $\alpha H(S_i) + H(\alpha, 1-\alpha) = H(S_{i+1})$. Subtracting $H(S_i)$ from both sides yields equation 2:

$$H(S_{i+1}) - H(S_i) = H(\alpha, 1-\alpha) - (1-\alpha)H(S_i) \quad (2)$$

We now show that this equation must be positive or zero to prove our theorem. If $\alpha = 1$ this obviously holds. Otherwise, the right side of the equation 2 is minimized with the maximum value of $H(S_i)$, which is $\log i$ and is obtained with all equal probabilities. Thus, we now consider equation 3.

$$H(\alpha, 1-\alpha) - (1-\alpha)H(S_i) \geq H(\alpha, 1-\alpha) - (1-\alpha)\log i \quad (3)$$

Since $f(\alpha) = H(\alpha, 1-\alpha) - (1-\alpha)\log i$ is a monotonically increasing function and $\alpha \geq \frac{i}{i+1}$, its minimum is obtained at $f(\frac{i}{i+1}) = (i+1)\{\log(i+1) - \log(i)\} \geq 0$. Therefore, $H(S_i) \leq H(S_{i+1})$ and by induction $H(S_i) \leq H(S_n)$ holds.

□

B. REFERENCES

[1] TeleNav. http://www.telenav.net/, 2004.

[2] Inrix. http://www.inrix.com/, 2006.

[3] Intellione. http://www.intellione.com/, 2006.

[4] D. Agrawal and C. C. Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In *Symposium on Principles of Database Systems*, 2001.

[5] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proc. of the ACM SIGMOD Conference on Management of Data*, pages 439–450. ACM Press, May 2000.

[6] A. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[7] A. Beresford and F. Stajano. Mix zones: User privacy in location-aware services. In *IEEE PerSec*, 2004.

[8] C. Bettini, X. SeanWang, and S. Jajodia. Protecting privacy against location-based personal identification,. In *2nd VLDB Workshop SDM*, 2005.

[9] R. Cayford and T. Johnson. Operational parameters affecting use of anonymous cell phone tracking for generating traffic information. *Institute of transportation studies for the 82th TRB Annual Meeting*, 1(3):03–3865, Jan 2003.

[10] D. Chaum. Untraceable electronic, mail return addresses, and digital pseudonyms. *Communications of the ACM*, 1981.

[11] A. Civilis and S. Pakalnis. Techniques for efficient road-network-based tracking of moving objects. *IEEE TKDE*, 17(5):698–712, 2005. Senior Member-Christian S. Jensen.

[12] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[13] L. Cranor, M. Langheinrich, M. Marchiori, and J. Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification. W3C Recommendation, Apr. 2002.

[14] X. Dai, M. Ferman, and R. Roesser. A simulation evaluation of a real-time traffic information system using probe vehicles. In *Proceedings of the IEEE Intelligent Transportation Systems*, pages 475–480, 2003.

[15] J. Deng, R. Han, and S. Mishra. Countermeasures against traffic analysis attacks in wireless sensor networks. In *Proceedings of the IEEE/Create-Net SecureComm*, Athens, Greece, September 2005.

[16] C. Diaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In *2nd Workshop on Privacy Enhancing Technologies*, 2002.

[17] R. Dingledine, N. Mathewson, and P. F. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, pages 303–320, 2004.

[18] A. Escudero-Pascual, T. Holleboom, and S. Fischer-Hubner. Privacy of location data in mobile networks. In *Proceedings of the 7th Nordic Workshop on Secure IT Systems (Nordsec 2002)*, 2002.

[19] H. Federrath, A. Jerichow, and A. Pfitzmann. Mixes in mobile communication systems: Location management with privacy. In *Proceedings of the First International Workshop on Information Hiding*, pages 121–135, London, UK, 1996. Springer-Verlag.

[20] M. Ferman, D. Blumenfeld, and X. Dai. A simple analytical model of a probe-based traffic information system. In *Proceedings of the IEEE Intelligent Transportation Systems*, pages 263–268, 2003.

[21] A. Gal and V. Atluri. An authorization model for temporal data. In *Proceedings of the 7th ACM CCS*, pages 144–153, New York, NY, USA, 2000. ACM Press.

[22] B. Gedik and L. Liu. Location privacy in mobile systems: A personalized anonymization model. In *Proceedings of the 25th IEEE ICDCS 2005*, pages 620–629, Washington, DC, USA, 2005.

[23] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. *Communications of the ACM (USA)*, 42(2):39–41, 1999.

[24] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the ACM MobiSys*, 2003.

[25] M. Gruteser and D. Grunwald. Enhancing location privacy in wireless lan through disposable interface identifiers: a quantitative analysis. In *Proceedings of the 1st ACM WMASH*, pages 46–55. ACM Press, 2003.

[26] M. Gruteser and B. Hoh. On the anonymity of periodic location samples. In *Proceedings of the Second International Conference on Security in Pervasive Computing*, 2005.

[27] B. Hoh and M. Gruteser. Protecting location privacy through path confusion. In *Proceedings of IEEE/Create-Net SecureComm*, Athens, Greece, September 2005.

[28] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[29] Y.-C. Hu and H. J. Wang. Location privacy in wireless networks. In *Proceedings of the ACM SIGCOMM Asia Workshop 2005*, April 2005.

[30] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. K. Miu, E. Shih, H. Balakrishnan, and S. Madden. CarTel: A Distributed Mobile Sensor Computing System. In *4th ACM SenSys*, Boulder, CO, November 2006.

[31] T. Jiang, H. Wang, and Y.-C. Hu. Preserving location privacy in wireless lans. In *Proceedings of the 5th ACM MobiSys*, New York, NY, USA, 2007. ACM Press.

[32] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk. Enhancing source-location privacy in sensor network routing. In *Proceedings of the 25th IEEE ICDCS'05*, pages 599–608, Washington, DC, USA, 2005.

[33] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar. Random data perturbation techniques and privacy preserving data mining. In *IEEE ICDM*. IEEE Press, 2003.

[34] J. Krumm. Inference attacks on location tracks. In *Proceedings of the Pervasive 2007*, May 2007.

[35] J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. In *Ubicomp*, pages 243–260, 2006.

[36] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran. Swing & swap: user-centric approaches towards maximizing location privacy. In *Proceedings of the 5th ACM WPES '06*, pages 19–28, New York, NY, USA, 2006. ACM Press.

[37] M. F. Mokbel, C.-Y. Chow, and W. G. Aref. The new casper: query processing for location services without compromising privacy. In *Proceedings of the 32nd VLDB'2006*, pages 763–774. VLDB Endowment, 2006.

[38] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. In *Proceedings of IEEE Symposium on Research in Security and Privacy*, 1998.

[39] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki. Caravan: Providing location privacy for vanet. In *3rd workshop on Embedded Security in Cars (ESCAR2005)*, 2005.

[40] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *2nd Workshop on Privacy Enhancing Technologies*, 2002.

[41] E. Snekkenes. Concepts for personal location privacy policies. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 48–57, New York, NY, USA, 2001. ACM Press.

[42] L. Sweeney. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588, 2002.

[43] K. P. Tang, P. Keyani, J. Fogarty, and J. I. Hong. Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications. In *Proceedings of CHI '06*, pages 93–102, 2006.

[44] J. M. Wozencraft and I. M. Jacobs. *Principles of Communications Engineering*. John Wiley & Sons Inc, 1966.

[45] M. Youssef, V. Atluri, and N. R. Adam. Preserving mobile customer privacy: an access control system for moving objects and customer profiles. In *Proceedings of the 6th MDM '05*, pages 67–76, New York, NY, USA, 2005. ACM Press.