# On the Anonymity of Periodic Location Samples

Marco Gruteser and Baik Hoh

Winlab / Electrical and Computer Engineering Department
Rutgers, The State University of New Jersey
94 Brett Rd
Piscataway, NJ 08854
{gruteser,baikhoh}@winlab.rutgers.edu

**Abstract.** As Global Positioning System (GPS) receivers become a common feature in cell phones, personal digital assistants, and automobiles, there is a growing interest in tracking larger user populations, rather than individual users. Unfortunately, anonymous location samples do not fully solve the privacy problem. An adversary could link multiple samples (i.e., follow the footsteps) to accumulate path information and eventually identify a user.
This paper reports on our ongoing work to analyze privacy risks in such applications. We observe that linking anonymous location samples is related to the data association problem in tracking systems. We then propose to use such tracking algorithms to characterize the level of privacy and to derive disclosure control algorithms.

## 1 Introduction

The continuous improvements in accuracy and cost of Global Positioning System (GPS) receivers are driving new location tracking applications with a massive user base. For example, in the United States, cell phone providers can determine the positions of emergency callers through Assisted GPS, and the German federal government is funding the development of a GPS-based highway toll collection system for trucks. These systems are capable of sampling location information from a large numbers of users.

We anticipate great demand for this data, going far beyond the original applications of emergency positioning or toll collection. Aside from hotly debated uses such as in law enforcement and targeted marketing, there are also clearly benevolent uses. For example, vehicles could report the location of abrupt braking activity to improve road safety, navigation systems that optimize traffic flows could alleviate congestion and pollution, or movement models collected from cell phones may help predicting the spread of infectious diseases.

Sharing location information, however, raises privacy concerns [1, 2]. For example, frequent visits to clinics signal medical problems, attending meetings may reveal political preferences, and meetings of influential business managers could indicate pending business deals. As such, the problem of sharing location information is analogous to hospitals publishing medical records to epidemiologists and other medical researchers—it can be beneficial to society but invades on privacy.

Anonymizing data provides a solution that enables data access while maintaining privacy. Sweeney [3, 4] pointed out, however, that naive anonymization strategies, such

as omitting names and street addresses, can in many cases be circumvented by a determined adversary. The combination of several factors (e.g., age, gender, zip code, race) may be sufficiently distinctive to correlate the data with other databases to reidentify individuals.

Similarly, Beresford and Stajano have reported in their pioneering work on the anonymity of location *traces* [5] how such traces can be identified. They then proposed the mix zone concept as an approach to split paths into unlinkable segments to increase privacy. In our earlier work [6], we have concentrated on the anonymity of *point* information, as used by many location-based queries. We developed mechanisms to dynamically adjust the accuracy of position information in location-based queries to maintain a predefined level of anonymity. These mechanisms were based on the assumption that queries are very sporadic and therefore can be adjusted independently. If sample points are revealed more frequently, the trajectory of a user may be used to link multiple samples and independence is not guaranteed. In this sense, time-series information like location differs significantly from medical records.

The class of applications considered here lies in between these concepts. Users report their location more frequently, so that the data cannot be anonymized as individual points, but they do not reveal a pseudonym that would link the points into location traces.

As such, this ongoing work can be viewed as bridging the gap between point anonymity and trace anonymity. We study how an adversary can exploit trajectory information to link anonymous location samples to location traces and identify multi-target tracking algorithms as a key threat. Based on these results we discuss the effect of sample rate on privacy and how formulations in multiple hypothesis tracking are helpful for deriving privacy mechanisms.

The remainder of this paper is structured as follows. Section 2 defines the class of applications and the privacy problem that this paper addresses. We introduce multi-target tracking algorithms in Sec. 3. Section 4 describes our experiments with such an algorithm on location samples collected through GPS and Sec. 4.2 discusses the results. Section 5 reviews related work before we conclude with Sec. 6.

## 2 Threat Assessment

We motivate the class of applications considered in this paper with an example from the automotive industry. There is interest in inferring traffic conditions from data collected in vehicles [7]. Selected vehicles could periodically send their location, speed, road temperature, windshield wiper status, and other information to a traffic monitoring facility. This data reveals the length of traffic jams (through speed and position), weather conditions such as rain (through windshield wiper activity), and slick road conditions (through frequent anti-lock braking). Using vehicles as mobile sensing platforms promises dramatic cost reductions over deploying specialized roadside sensors.

Generally, we will consider a class of remote data collection applications. This class of applications requires a large number of users to concurrently reveal anonymous location information to an external untrusted service provider. The data are collected with a well-known sample frequency $f$. We can characterize the data that an external service

provider receives as follows. The data comprises a series of tuples containing sensor data, latitude, longitude, and time. The sensor data could be any sensor reading associated with this location such as road temperature or anti-lock braking activity. We assume, however, that the sensor readings themselves do not contain any information distinctive enough to enable tracking of individual users. Latitude and longitude can be initially determined locally on a user's device (e.g., in a car), through external infrastructure (e.g., a cell phone provider's network), or hybrid approaches. We assume, however, that the user can trust the device or infrastructure that initially senses position.

We also assume the existence of a trusted proxy that anonymizes location updates before they are passed on to the external service provider. In a cell phone based system, for example, the proxy could arguably be operated by the cell-phone provider, who operates the location tracking infrastructure and sends data to third-party application providers. Anonymizing location updates means removing identifier like user ids or network addresses, but also mixing of messages to counter timing attacks. Furthermore, this means that we will only consider applications that do not depend on user identity information.

### 2.1 Inference Attacks

In this paper, we will concentrate on inference attacks based on the data that the external service provider received. We will not consider attacks against the infrastructure that determines, transmits, or processes a user's location—we assume it has been appropriately secured. The inference attacks may be carried out by the service provider, malicious employees of this provider, or by anybody else who has legitimately or illegitimately gained access to this information. We are most concerned, however, with attacks that can be easily automated to monitor larger groups of individuals. We will not consider how this data could be used in targeted investigations against a specific individual.

This class of applications at first does not appear to bear any privacy risks, because each tuple is revealed anonymously. On second thought, however, it becomes clear that an adversary could link independent updates to the same user if the sample frequency $f$ is sufficiently high compared to the user density in an area. This leads to an accumulation of path information about individual users that will likely lead to identification. For example, Beresford and Stajano [5] report that the location traces collected in an office environment through the Active Bat system could be correctly reidentified by knowing the desk positions of all workers and correlating them with the traces.

Informally, the privacy property that this research aims for is unlinkability of location samples. An adversary could employ at least three approaches to link location samples. First, trajectory-based linking assumes that a user is more likely to continue traveling on the same trajectory, rather than changing direction. The adversary could build a basic movement model that includes probabilities for altering a course from a sample user population. Second, map-based linking correlates location samples with likely routes on a road or building map. The routes can then be used to predict users' position and to link future samples. Third, empirical linking connects samples based on prior movements that have been observed at a given location.

We believe that trajectory-based linking requires the least effort for large-scale outdoor positioning systems. The adversary does not have to gather map information or collect empirical information for every intersection. Therefore, we will restrict our analysis on this approach.

## 3 Multi Target Tracking

The tracking systems community knows the problem of linking location samples to probable users as the data association problem in multi-target tracking systems. Radar provides one typical application: the system must assign anonymous radar echos to a set of tracked targets. The key idea of such algorithms is to compare the positions of new location samples with the predicted positions of all known targets and choose an assignment that minimizes the error.

We chose Reid's multiple hypothesis tracking algorithm [8], which is based on Kalman filtering. This algorithm is one of the basic works in the field [9, p. 325]. Although, we do not currently use its capability to maintain multiple hypotheses, we have chosen it because we plan to experiment with this feature in future work.

Here, we will summarize our implementation of the algorithm. We refer the reader to the original work [8] for a more in depth discussion and the derivation of the equations. Additional information, also on the Kalman filter, can be found in [9]. The algorithm operates in three steps: First it predicts a new system state, then generates hypotheses for the assignment of new samples to targets and selects the most likely hypotheses, and finally it adjusts the system state with information from the new samples.

We simplified Reid's algorithm in a number of points. First, we do not consider random track initiation. Second, we assume all samples are taken at a fixed sample rate. Finally, as already mentioned, after every step only one hypothesis survives, which means that at each step likelihood is calculated under the assumption that the previous assignments were correct.

### 3.1 State Prediction

The filter predicts state according to a process model that is described by

$$x_k = Fx_{k-1} + w,$$

where $x_k$ is the state vector of the process at step $k$, matrix $F$ describes a linear prediction of the next state given the previous state, and $w$ represents the process noise vector. A new observation vector $z_k$ relates to the actual state through

$$z_k = Hx_k + v,$$

where matrix $H$ converts a state vector into the measurement domain and $v$ represents the measurement noise vector. The filter assumes that the process noise and the measurement noise are independent of each other and normally distributed with covariance matrices $Q$ and $R$, respectively.

When tracking only one target, the Kalman filter defines the conditional probability density function of the state vector at time instant $k$ as a multivariate normal distribution with mean $\bar{x}$ and covariance $\bar{P}$. At each time step, the filter predicts the new target position as

$$\bar{x}^{k+1} = F\hat{x}^k \qquad \text{and} \qquad \bar{P}^{k+1} = F\hat{P}^k F^T + Q^T, \tag{1}$$

where $\hat{x}$ and $\hat{P}$ are the estimates after the last sample was received.

For two-dimensional tracking applications with only slight changes in trajectory we can model the system as

$$F = \begin{bmatrix} 1\,0\,1\,0 \\ 0\,1\,0\,1 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix} \qquad x = \begin{bmatrix} p_x \\ p_y \\ v_x \\ v_y \end{bmatrix},$$

where $(p_x, p_y)$ represent a position and $(v_x, v_y)$ a velocity vector. A larger process noise component captures the probability of changing directions or velocity.

### 3.2 Hypotheses Generation and Selection

The algorithm generates a set of hypotheses when new samples are received—one for each permutation of the sample set. A hypothesis represents a possible assignment of new samples to targets. It then calculates the likelihood for each hypothesis and selects the one with maximum likelihood.

The probability of hypothesis $\Omega_i$ at time $k$, given the set of measurements $Z^k$ with cardinality $M$, is described by

$$P_i^k \equiv P(\Omega_i^k \mid Z^k) \approx \prod_{m=1}^{M} f(z_m) \tag{2}$$

where $f$ is defined by the following equation (3). Based on the observation equation in the Kalman filter, the conditional probability density function of the observation vector $z_k$ obeys a multivariate normal distribution

$$f(z^k \mid \bar{x}^k) = N(z^k - H\bar{x}^k, B), \tag{3}$$

where $B = H\bar{P}^k H^T + R$ and $N(x, P)$ denotes the normal distribution

$$N(x, P) = e^{-\frac{1}{2}x^T P^{-1} x} / \sqrt{(2\pi)^n \mid P \mid}.$$

Both $x^k$ and $P$ are calculated using the update equation at the prediction step. Equation (3) calculates how close a new observation lies to a predicted position; these values are then combined into the probability of each hypothesis.

After calculating the probability of each hypothesis, we choose the hypothesis $j$ with the maximum probability and also calculate the log-likelihood ratio as follows.

$$\log \Lambda^k = \log \frac{P_i^k}{\sum_{i=1, i\neq j}^{I} P_i^k} \tag{4}$$

### 3.3 State Correction

In the correction step, the predicted system state vector for each path will be updated with the Kalman gain and the difference between the assigned observation vector and the predicted vector. The observation vectors are assigned to the targets according to the chosen hypothesis. Equation (5), which is similar to a recursive least square update equation, describes the correction step. In this equation $K = \hat{P}H^T R^{-1}$ is the Kalman gain.

$$\hat{x}^k = \bar{x}^k + K[z^k - H\bar{x}^k] \qquad \text{and} \qquad \hat{P}^k = \bar{P} - \bar{P}H^T(H\bar{P}H^T + R)^{-1}H\bar{P} \quad (5)$$

The so corrected state vector and covariance matrix are then fed back into the prediction equations and the steps are repeated for the next set of samples.

## 4 Experimentation

To evaluate the privacy risks posed by multi-target tracking, we have asked students on a university campus to carry an off-the-shelf GPS receiver as they go about their daily routine. Here we discuss our preliminary results in applying multi target tracking algorithms to a first batch of data. We also present micro-benchmarks that illustrate how multi-target tracking computes the relative likelihood of an assignment.

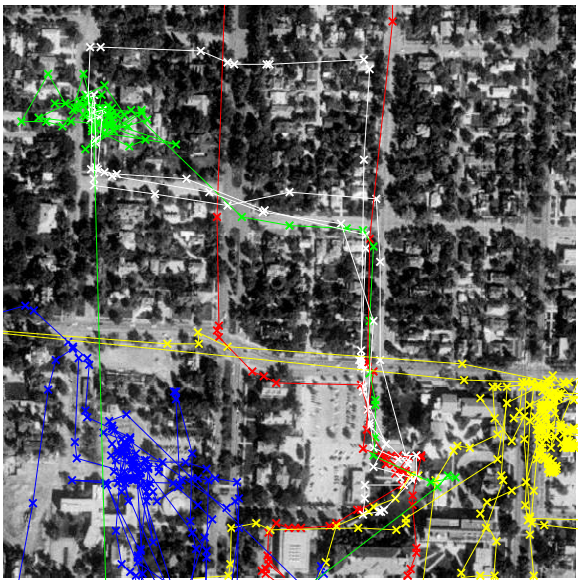

**Fig. 1.** Five GPS paths shown over a satellite image of the area. The paths contain several clusters, where users stayed for an extended time. There are also several areas where different users' paths overlap.

The first batch of GPS data comprises five sample tracks of students, each collected over the course of one day. Figure 1 shows their tracks plotted onto aerial imagery. The tracks intersect on campus, but also extend off-campus. It includes a mix of pedestrian and vehicle movements. In short, this data provides us with a sample of students' movement patterns. It is not intended to allow drawing conclusions about user density or 'average' mobility of a larger population.

Unfortunately, two of the paths were rather short, so we decided to run our first experiments only on the three longer ones (depicted in Fig. 2). We also chose two micro-benchmarks: two nearly orthogonal path segments that intersect in space and two parallel segments that first merge and later diverge. Both cases represent extreme inputs to a two target trajectory-based tracking system. The system should perform best on the orthogonal intersection; the parallel scenario should prove most challenging.

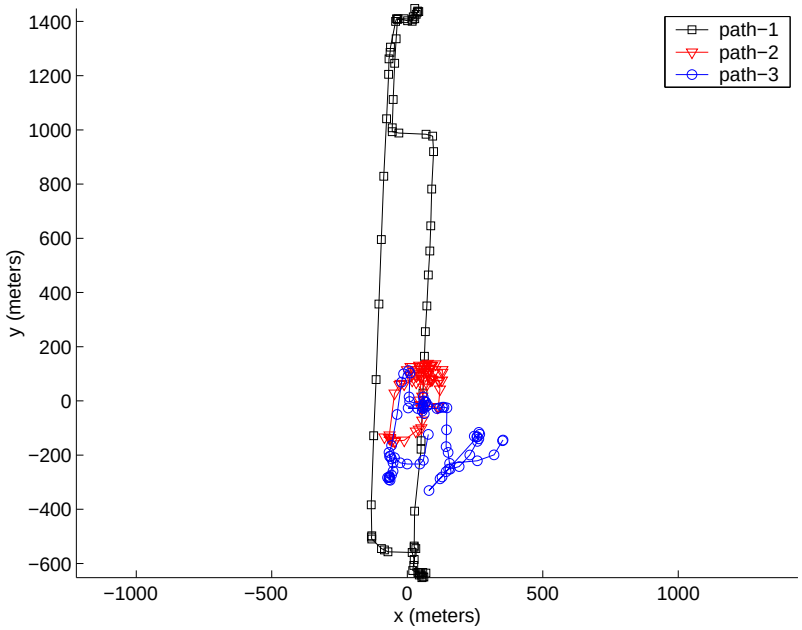


**Fig. 2.** Three longer paths used for the target tracking experiment.

Figures 3 and 4 show the chosen path segments. For the experiment, we removed the labels that identify to which path a given sample belongs. At each step in the tracking process we supply the next two samples and let Reid's algorithm solve the assignment problem. Note that in the actual paths, the two users did not pass this area on the same day, therefore they can be trivially distinguished based on the timestamps. To make this scenario challenging, we have adjusted the timestamps so that the two users simul-

taneously arrive at the intersection or simultaneously start out on the parallel tracks, respectively.

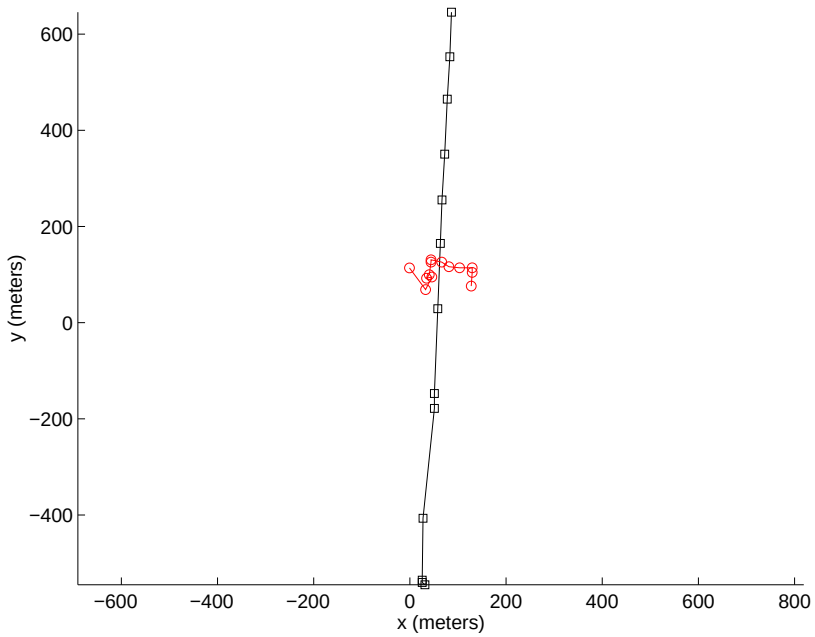


**Fig. 3.** Two orthogonally intersecting path segments extracted from the GPS data. One user is moving north, the other east.

Reid's MHT algorithm depends on several parameters that affect its performance. We refined the process model described in Sec. 3 by applying an expectation maximization algorithm [10] that estimates the parameters based on the five location tracks. We implemented both Reid's algorithm and the EM algorithm in MATLAB. To simplify the implementation, we first converted the GPS data into the Universal Transverse Mercator System projection, where (within one zone) a position is described in meters on a cartesian coordinate system.

### 4.1 Results

Figure 5 describes the result of applying the multi-target tracking algorithm to the three longer paths. The three curves show the paths that the algorithm reconstructed from the anonymous samples. A change in value of a curve means that the algorithm has misassigned samples—three constant curves would mean perfect reconstruction. The algorithm clearly confuses a number of sample points, but many misassignments are only temporary. The first path is correctly tracked until sample 52, the second path has more misassignments, but recovers and is correctly assigned at the end. Only the third path exhibits sufficient track confusion to provide a high level of privacy.

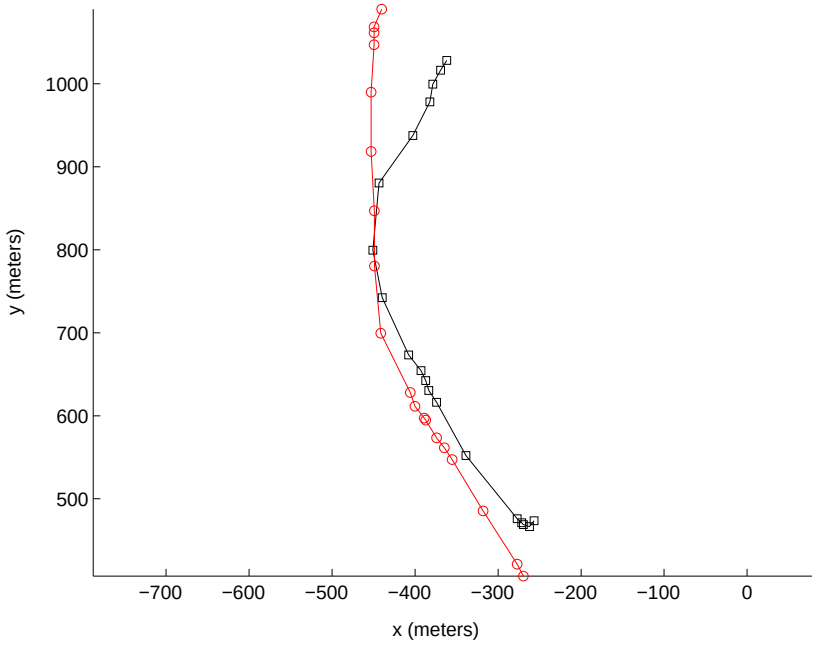


**Fig. 4.** Two parallel path segments extracted from the GPS data. Both users move north.

Figure 6 shows the log-likelihood ratio at each step in the MHT process for the orthogonal path segments. The log-likelihood ratio is computed as described in Sec. 3.2. Thus, higher positive values indicate more confidence in the assignment, a value of zero corresponds to equally likely hypotheses, and a negative value to a false assignment.

As shown in the curve with square points, assignment certainty decreases as the two targets approach each other. Their paths intersect at sample index 8, where the log-likelihood ratio actually dips below zero; these particular samples are falsely assigned. The algorithm recovers, however, with the following sample updates. At sample index 10, after the paths diverged, paths are assigned with high confidence again. This example illustrates how the tracking algorithm disambiguates two intersecting paths. The curve with round points depicts log-likelihood for the same scenario but with only half the sampling rate. We see that reducing the sampling rate results in only small changes in log-likelihood for the first few samples. After the intersection, however, the algorithm now tracks the wrong path.

The curve with round points in Fig. 7 shows the log-likelihood graph for the parallel segments. For these paths, there is not much information that allows the algorithm to distinguish them. The algorithm falsely assigns samples 3–6. Note that the confidence for the fifth sample is comparatively high, even though it is misassigned. This illustrates how an error can propagate. The next point of confusion starts at sample 6, where the log-likelihood again approaches zero.

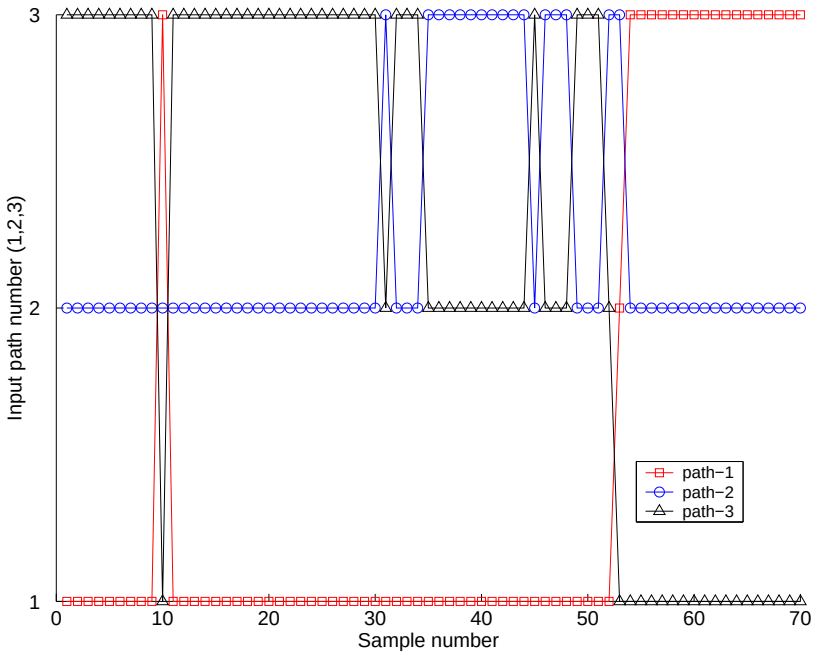


**Fig. 5.** Disambiguation of paths. The three curves represent the output paths. Each point of a curve shows from which input path the sample was taken. In short, where the curves cross the algorithm misassigned samples.

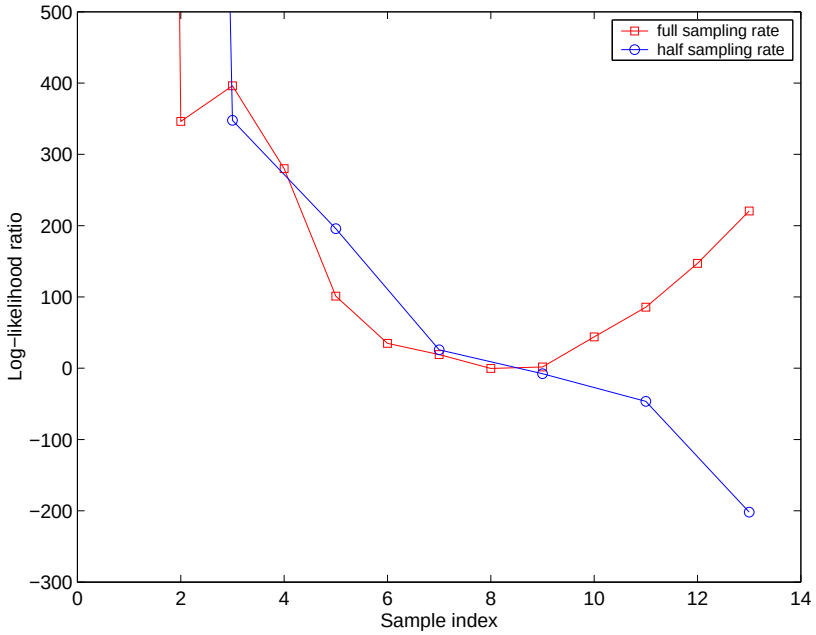


**Fig. 6.** Log-likelihood ratio at each step in the MHT tracking process for the orthogonal paths.

The curve with square points illustrates the sensitivity of the outcome with regard to slight changes in the movement model. This time we generated the model using only the two path segments rather than the complete five graphs. With this change the association outcome is now reversed. Only the first few points are correctly assigned.

### 4.2 Discussion

Note that log-likelihood ratios close to zero are neither a necessary nor a sufficient condition for location privacy. It is not sufficient because location privacy depends on how accurate an adversary can estimate a user's position. An adversary may not be able to disambiguate the samples from two users, but if the samples are very close to each other, the location estimate for both users is still relatively accurate. The parallel paths scenario illustrate this case. The adversary may misassign the samples, but this leads at most to a difference of about 100m. A low log-likelihood ratio is also not a necessary condition because in some cases the adversary will assign samples with high likelihood, but the assignment can still be wrong. This can be observed in Fig. 7, where the log-likelihood falls below -50. If the sample rate was lower, we could observe such low values without first encountering values close to zero.

We can, however, use dropping log-likelihood ratios as an indicator of potential track confusion. Location privacy increases, when the adversaries tracking algorithm confuses paths from at least two users and these paths then significantly diverge.

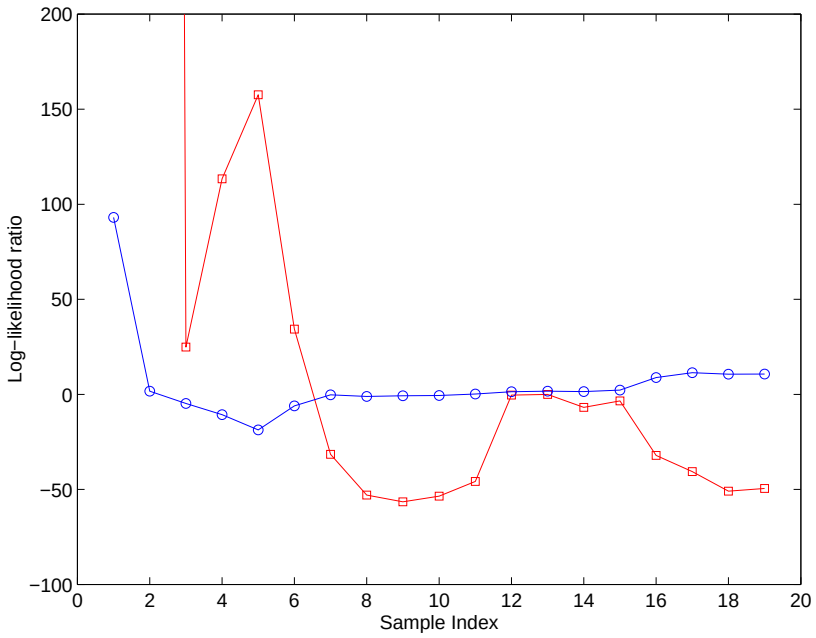


**Fig. 7.** Log-likelihood ratio at each step in the MHT tracking process for the parallel paths. The curve with square points shows the results with a different movement model. Tracking performance is very sensitive to slight changes in the movement model.

In our experiments, we have used a linear model with a large white gaussian noise component to model nonlinear user movements. We have seen that this crude model could already provide useful tracking results to an adversary. Tracking results could probably be improved by using nonlinear filters such as Extended Kalman Filtering and colored noise models, where the amount of noise at a given step depends on the previous step.

## 5 Related Work

The presented results build on prior work in the field of location anonymity. We have analyzed a new class of applications that requires periodic location samples from a larger number of users to infer statistical properties (such as average traffic flow).

In prior work [6], we have described mechanisms to guarantee a defined degree of anonymity in different locations by adjusting the spatio-temporal resolution of location-based queries. These mechanisms assume that location-based queries are generated so infrequently, that they can be viewed as independent queries—the adversary would be unable to link them to the same user. The current paper has described a first analysis aimed at building mechanisms to detect when the frequency of queries becomes dangerously. Another goal of these mechanisms could be to control this frequency so that a defined level of privacy is maintained.

This research is also closely related to the mix zone concept developed by Beresford and Stajano [5, 11]. Mix zones are spatial areas in which users' location is not accessible. When multiple users simultaneously traverse a mix zone, their pseudonyms can be changed and it becomes hard to link the incoming and outgoing path segments to the same user. In this way mix zones can be viewed as segmenting paths. They are more suitable for applications that require quasi-continuous tracking of users during a particular time interval, rather than the less frequent location samples that we discussed in this paper. However, we believe that the multi-target tracking concepts will also be helpful in analyzing unlinkability of paths over mix zones.

Another thread of privacy research for location-aware systems [12–14] develops privacy policy-based technologies to make users aware of a service provider's data collection practices. It also allows them to easily express preferences that govern under what circumstances private data can be shared. Lederer and colleagues [15] found that the identity of the requester typically is the most significant factor in users' privacy decisions. These mechanisms allow sharing information with trusted parties, while blocking intrusions from untrusted ones. Our location anonymity research is orthogonal to this work. To our knowledge privacy legislation does not mandate data collectors to inform users about *anonymous* data collection. As discussed, however, anonymity is not an absolute property, rather data can afford different degrees of anonymity. Therefore, privacy-policy mechanisms could be used to negotiate an acceptable degree of anonymity between users and service providers.

Serjantov and Danezis [16] as well as Diaz and colleagues [17] have proposed an information theoretic metric for anonymity. The metric was presented in the context of anonymous network communication but appears also applicable to location information. A privacy criterion for periodic samples will likely build on this work.

Privacy-aware data-mining follows a similar objective in allowing inferences about aggregate distributions of users while preserving privacy [18]. It differs in that it does not attempt to maintain anonymity, but rather protect sensitive data about users. For example, one mechanism perturbs sensitive data, such as salary information, by adding a random offset. This hides an individual user's information within an uncertainty interval, while still allowing the reconstruction of the salary distribution for a large user population. These mechanisms also do not address how to perturb time-series data such as location traces.

## 6 Conclusions

In this paper, we have considered a class of applications that requires a large number of users to reveal periodic location samples. This class is not yet adequately addressed by existing location privacy mechanisms. We have analyzed how multi-target tracking algorithms reconstruct paths from periodic anonymous location samples and proposed to derive a privacy criterion and disclosure control algorithms based on the inherent uncertainty metrics.

From our experiments, we have obtained the following insights. First, while the tracking performance of our implementation was not perfect, it did track the three users for an extended period of time. Most of the confusion between users is only temporary, when two paths cross, and not significant in the long run. Second, reducing the sampling rate with which location samples are published does not have a major effect on the certainty of assignment, unless it coincides with changes in direction or intersecting paths. Third, spoint-wise log-likelihood as a measure of uncertainty is not a good indicator of privacy per se. A path certainty measure that takes into account the uncertainty at previous sample points may be a better alternative. Log-likelihood appears to be a good predictor of potential confusion, though.

We see three important directions for continuing this work. First, we plan to develop an anonymity criterion that signals whether the sampling rate and user density parameters in a given application scenario meet a defined level of anonymity. This criterion should be guided by the performance of more refined versions of the tracking algorithm. In particular, we plan to study the effect of track initiations and of maintaining multiple likely hypotheses over a number of steps.

Second, we are interested in deriving disclosure control algorithms that could dynamically adjust the sampling rate to meet a privacy criterion. As discussed, reducing the sampling rate is most effective when it coincides with unpredictable changes in trajectory. Compared to a static privacy-preserving sampling rate, may provide an overall higher data quality to applications by only reducing the rate when needed. A full solution must also take location-based services' data quality requirements into account. Eventually, it should remain the choice of users and system designers to decide when to trade privacy for reduced service quality.

Eventually, such privacy mechanisms must be compared in light of application's data requirements. Privacy can be trivially improved by reducing the amount of data available, but this may not be adequate for a given application. Once we developed a better understanding of how to define a privacy criterion we also plan to clearly define

the data requirements for different applications. The most interesting problem will be to find algorithms that maximize privacy while maintaining the required data quality.

## Acknowledgments

We thank Jonathan Bredin and his students for fruitful discussions and for providing us with the GPS traces.

## References

1. Jay Warrior, Eric McHenry, and Kenneth McGee. They know where you are. *IEEE Spectrum*, Jul 2003.
2. Louise Barkhuus and Anind Dey. Location-based services for mobile telephony: a study of users' privacy concerns. In *9th Internation Conference on Human-Computer Interaction (INTERACT)*, 2003.
3. Latanya Sweeney. $k$-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.
4. Latanya Sweeney. Achieving $k$-Anonymity Privacy Protection Using Generalization and Suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588, 2002.
5. Alastair Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
6. Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the First International Conference on Mobile Systems, Applications, and Services*, 2003.
7. Rajiv Vyas. Ford device intended to unclog roads. `http://www.freep.com/money/autonews/ford27_20040227.htm`, Feb 2004.
8. Donald Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843–854, Dec 1979.
9. Samuel Blackman and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
10. Todd Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60, Nov 1996.
11. Alastair Beresford and Frank Stajano. Mix zones: User privacy in location-aware services. In *IEEE Workshop on Pervasive Computing and Communication Security (PerSec)*, 2004.
12. Ginger Myles, Adrian Friday, and Nigel Davies. Preserving privacy in environments with location-based applications. *IEEE Pervasive Computing*, 2(1):56–64, 2003.
13. Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In *4th International Conference on Ubiquitous Computing*, 2002.
14. Sastry Duri, Marco Gruteser, Xuan Liu, Paul Moskowitz, Ronald Perez, Moninder Singh, and Jung-Mu Tang. Framework for security and privacy in automotive telematics. In *2nd ACM International Worksphop on Mobile Commerce*, 2002.
15. Scott Lederer, Jennifer Mankoff, and Anind Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI)*, pages 724–725, 2003.
16. Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *2nd Workshop on Privacy Enhancing Technologies*, 2002.

17. Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *2nd Workshop on Privacy Enhancing Technologies*, 2002.
18. Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In *Proc. of the ACM SIGMOD Conference on Management of Data*, pages 439–450. ACM Press, May 2000.