Abstract—The prevalence of mobile devices with geopositioning capability has resulted in the rapid growth in the amount of moving object trajectories. These data have been collected and analyzed/mined for both commercial (e.g., recommendation system) and security (e.g., surveillance and monitoring system) purposes. It is clear that one needs to ensure the privacy of raw trajectory data by not disclosing or releasing the data to adversary. Moreover, existing trajectory pattern mining functionalities have the capability to derive knowledge from these raw data. In this paper, we assume that the trajectory data is secure and users can only query the knowledge derived from the trajectory data. We describe and discuss how one can ensure an individual’s privacy with respect to his location and spatiotemporal behavioral patterns via differential privacy mechanism that we proposed recently; in particular, we demonstrate the privacy preserving approach on the frequent location pattern mining task. Moreover, we discuss (i) the extension of its use to sequential pattern mining and association rules generation, and (ii) the dilemma of differential privacy goal for moving objects data mining tasks such as anomaly detection and outlier detection that are important techniques for terrorism related analytical approaches and crime forecasting.

I. INTRODUCTION

With the prevalence of location acquisition technology on GPS and mobile devices, there is an increasing trend in the collection of human and vehicular movement data for behavioral analysis, pattern mining, and knowledge extraction [1], [2]. These data are stored in moving objects databases [3], [4], [5] that extend database technology so that any kind of moving entity (e.g., moving point, line, or region data types) can be represented in a database and queries on the object movements can be formulated in a precise and yet simple query language extended from the conventional SQL. These enhanced database systems are full-fledged DBMSs with additional support for the management of spatial and moving objects data. In particular, one can perform data analysis and mining on the moving objects in the databases. To obtain meaningful patterns, a large and diverse amount of individual location history records has to be collected for such data analysis or mining. The privacy goal for this scenario is to ensure that an individual’s participation in such a moving objects statistical database does not substantially increase risk to his privacy. The desired privacy level for such a goal is captured and controlled by the measure of differential privacy [6] and has been demonstrated to work well for location data mining tasks [7]. With such privacy guarantee that limits risk when one’s location history is included into a moving objects database, an individual would be more willing to allow his location history to be collected. Moreover, with the increase number of participation, the moving objects database with pattern mining capabilities will then have higher utility.

Pattern mining from a moving objects statistical database enables one to extract useful location knowledge such as an ‘interesting location’ which can be understood as a geographic region that many people frequently visit. Other mining capabilities include sequential pattern mining and association rules generation. For a query on a moving objects statistical database with data mining capabilities, one assumes that only statistics and patterns are returned to users. Without disclosing a person’s membership in the moving objects database or even when the person’s location history is not in the database, an attack can take place to gain information about the person on the discovered location patterns. Hence, one main concern for individuals to participate in such data collection is the disclosure of their location and behavioral information when a user repeatedly perform queries that analyze the data with some auxiliary information [8].

In this paper, we assume that the trajectory data is secure and users can only query the knowledge derived from the trajectory data. We describe and discuss how one can ensure an individual’s privacy with respect to his location and spatiotemporal behavioral patterns via our recently proposed differential privacy mechanism; in particular, we demonstrate the privacy preserving approach on the frequent location pattern mining task. Moreover, we discuss (i) the extension of its use to sequential pattern mining and association rule generation, and (ii) the dilemma of differential privacy goal for mining tasks such as anomaly detection and outlier detection that are important techniques for terrorism related analytical approaches and crime forecasting.

The paper is organized as follows. In Section II, we describe previous work on privacy preserving data mining and privacy solutions for spatial-temporal data domains. In Section III, we discuss the privacy issues related to the three mining tasks using Semantic Trajectory Data Mining Query Language (STDMQL) [9]. In Section IV and V, we provide background knowledge for differential privacy and our proposed differential privacy mechanism demonstrated using frequent location pattern mining, respectively. In Section VI, we discuss how our proposed differential privacy mechanism can be extended to
the sequential pattern mining and association rule generation, and the difficulty of applying differential privacy mechanism on anomaly and outlier detection.

II. REVIEW

Conventional privacy goals such as data/response perturbation, k-anonymity [10], and differential privacy [6] have been considered for privacy preserving data mining algorithms for statistical databases. An algorithm achieves privacy protection if it can still perform well (i.e., accurate prediction model) when the records in the database are perturbed [11]. Although such a privacy goal is intuitive from a data mining perspective, it lacks theoretical results to establish a relationship between the model accuracy and the amount of data/response perturbation. An algorithm achieves k-anonymity privacy protection if each record in the database cannot be distinguished from at least \( k - 1 \) records in the database when the algorithm is applied [10]. An algorithm achieves \( \epsilon \)-differential privacy protection if the outputs returned from the algorithm for any two databases differ on a single record are approximately the same described by a probabilistic bound using an exponential function of \( \epsilon \) [6].

Privacy is an important issue for both spatial and spatiotemporal data mining due to the challenge of protecting personal location information [12], [13], [14] (and the references therein). k-Anonymity approaches are used frequently to preserve privacy during personal location collection in location-based services [15] or to perform privacy preserving spatial (range or nearest neighbors) queries [16]. Obfuscation or perturbation techniques are used to hide and confuse an adversary by modifying a user’s trajectory or location history. To achieve anonymity in publishing trajectory dataset is also a challenging problem. Monreale et al. [17] proposed a method to transform trajectory data based on spatial generalization (i.e., replacing exact locations by their approximations) and k-anonymity justified by a theoretical upper bound to the probability of re-identification. Nergiz et al. [18] proposed a method that ensures k-anonymity and included sampling from the anonymized data to prevent leakage due to the anonymization step. Yarovoy et al. [19] introduced a form of k-anonymity based on spatial generalization for moving objects and proposed two anonymization approaches that achieve their notion of k-anonymity. Recently, Abul et al. [20] also introduced the concept of \((k, \delta)\)-anonymity that exploits the inherit location uncertainty of moving objects, represented by radius \( \delta \), to reduce the amount of distortion needed to anonymize data using clustering and spatial perturbation.

There have been many research work on preserving privacy for sensitive information in statistical databases [11] (and the references therein) and privacy preserving knowledge discovery techniques [21] (and the references therein). Decision tree classifier is the most commonly used data mining algorithm that research has been done to explore privacy issues [11], [22], [23]. Agrawal and Srikant [11] addressed the task of developing accurate decision tree using perturbed data sets to preserve data record privacy. Lindell and Pinkas [22] developed theoretically a protocol on the task when two parties perform data mining, in particular running the ID3 decision tree classifier, on the union of their confidential databases without revealing any confidential information to each other. Friedman and Schuster [23] used the Privacy Integrated Queries (PINQ) API [24], an infrastructure for implementing differentially private algorithms, to construct a differentially private ID3 decision tree classifier.

Recently, there have been great interests in developing machine learning and data mining algorithms with differential privacy guarantees. McSherry and Mironov [25] adapted the leading machine learning approaches in the Netflix Prize competition to perform with differential privacy guarantees so that the learned correlations and an individual’s data can be used to provide personalized recommendations. Friedman and Schuster [23] investigated differentially private data mining algorithm focusing on the decision tree algorithm. Chaudhuri et al. [26] constructed differentially private regularized logistic regression and support vector machines by perturbing the objective functions for these machine learning algorithms.

Before we describe privacy issues for moving objects data mining in Section III, we briefly describe Semantic Trajectory Data Mining Query Language (ST-DMQL) [9] that we used to illustrate moving objects data mining queries. ST-DMQL is a data mining query language based on moving object trajectories and geographic information. Meaningful location patterns can be extracted from trajectory and geographic information. ST-DMQL has basic functionalities such as select, trajectory preprocessing, and data mining capabilities.

III. PRIVACY ISSUES IN MOVING OBJECTS DATA MINING

There are many data analysis functions that can be included into a moving objects database to extract knowledge from the moving objects data. These functions are used to handle data mining tasks such as frequent pattern discovery, sequential pattern discovery, and association rule generation. To control the level of differential privacy for these data analysis functions, one has to apply differential privacy mechanism (see Section IV) to the frequency or support count, \( \sigma(G) \), for a pattern/rule \( G \).

For simplicity and without loss of generality, let \( I = \{i_1, i_2, \ldots, i_d\} \) be the set of geographic location items containing information about both space and time defined in [9], [2]\(^1\) the following data types or structures.

- **Name**: stop (e.g. RestaurantA); move (e.g. Airport-Hotel);
- **NameStart**: stop/move with start time period (e.g., MuseumA[morning], HotelB-MuseumC[afternoon]);
- **NameEnd**: stop/move with end time period;
- **NameStartEnd**: stop/move with start/end time period (e.g., MuseumA[morning][afternoon])

Let \( traj_k^i = \{p_1, p_2, \ldots, p_k\} \) be a trajectory consisting of \( k \) measurements. Each \( p_j = (x_j, y_j, t_j) \) such that \( x_j \) is the latitude, \( y_j \) is the longitude, and \( t_j \) is a timestamp and \( t_j <

\(^1\)In this paper, we ignore the data preprocessing step. Interested readers can refer to [9], [2].
We define a stay point to be the center \((x, y)\) of a circle region that a trajectory stays for at least a time period of \(\Delta T\). The support count \(\sigma(G)\) is the number of stay points for a geographic location item. Without considering the temporal aspect of an event, a frequent \(k\)-itemset \(F = \{i_{f1}, \ldots, i_{fk}\} \subset I\) such that the support count for each item is greater than some user defined threshold. A sequential itemset \(S = (i_{s1}, \ldots, i_{sl})\) is an ordered sequence of items in \(I\) such that the support count considering their order and item frequencies is greater than some user defined threshold. Next, we describe three moving objects data mining tasks, namely: frequent pattern mining, sequential pattern mining, and association rule generation, that may disclose an individual’s privacy when his trajectory is added or removed from the database.

### A. Frequent Patterns

Using the convention in [9], we construct a query that can answer the question: “What are the most frequently visited recreational locations by people during weekend?” as follows.

```sql
SELECT frequentStops
    (item = NameStart, timeG = WEEKEND,
     stopG = [type, EntertainmentPlaceH = 1],
     minsup = 0.2)
FROM stop
with a possible query result:

{Cinema[weekend]} (s = 0.45)
{Restaurant[weekend]} (s = 0.30)
{RecreationPark[weekend]} (s = 0.20)
```

The query result is based on the user defined minimum support, \(minsup\) and \(s\) is the percentage of stay points in the query result, i.e., \(s = \frac{\sigma(G)}{t_{sp}}\), where \(t_{sp}\) is the total number of stay points. Note that this is different from [9] where \(s\) is the percentage of trajectory, i.e., as long as there is one stay point, a trajectory will be included in the count.

If an individual’s information is removed from the database, his favorite weekend entertainment may be disclosed based on a change in the support value, \(\sigma(G)\). For example, when an individual’s trajectory is removed from the moving objects database, \(\sigma(Cinema[weekend])\) drops from 450 to 400 (say for a one-year period). It may signify that the individual goes to cinema very regularly during weekends.

### B. Sequential Patterns

Similar to the frequent location pattern task, the sequential location pattern query “What are the most popular travel sequence of people during weekend?” can be constructed as follows.

```sql
SELECT sequentialMoves
    (item = NameEnd, timeG = WEEKEND,
     stopG = instance, minsup = 0.05)
FROM move
```

with a possible query result:

\{RestaurantA, CinemaB\} (s = 0.09)
\{CinemaB, RestaurantC\} (s = 0.07)

### C. Association Rules

To generate interesting relation between location(s) is also an important pattern mining task for moving objects. An associate rule query “What are the rules on places people like to go during weekends?” is constructed as follows.

```sql
SELECT associateStops
    (item = NameStart, timeG = WEEKEND,
     stopG = instance, minsup = 0.05,
     minconf = 0.40)
FROM stop
```

Applying the Apriori Algorithm [27] on location variables with user defined minimum confidence, \(minconf\), and minimum support, \(minsup\), the query result is of the following form.

\{RestaurantB, CinemaA\} \rightarrow \{ParkA\}
(s = 0.09) (c = 0.70)
\{CinemaB\} \rightarrow \{RestaurantC\}
(s = 0.07) (c = 0.40)

where \(c\) is the confidence level.

For all three mining tasks, when an individual’s trajectory is removed from or added into the database, the support, \(s\), needed to extract the pattern will change. As a result, one can infer from the change in support an individual’s frequent location, his/her travel behavior (order sequence), and his/her interesting travel characteristics (association rules).

### IV. Differential Privacy

Differential privacy ensures that one can extract meaningful knowledge or information from a database while privacy is preserved for any individual whether his data is in the database or not.

**Definition 1:** [8] A randomized function \(K\) gives \(\epsilon\)- differential privacy if for all databases \(D_1\) and \(D_2\) differing on at most one element, and all \(S \subset \text{Range}(K)\),

\[Pr[K(D_1) \in S] \leq \exp(\epsilon) \times Pr[K(D_2) \in S]\]

The function \(K\) satisfies the fact that when one element is removed from the database, no output would become significantly more or less likely. From another perspective, \(\epsilon\)-differential privacy for a pattern mining algorithm can be defined as follows.

**Definition 2:** A pattern mining algorithm \(M\) provides \(\epsilon\)-differential privacy if for any two databases \(D_1\) and \(D_2\) that differ in a single entry and for any \(a\),

\[
\log \frac{\mu(M(D_1) = a|D_1)}{\mu(M(D_2) = a|D_2)} \leq \epsilon
\]

such that \(M(D)\) is the random variable that represents the algorithm output and \(\mu\) denotes the probability density for the algorithm output. Inequality (1) is also called the \(\epsilon\)– indistinguishable and \(\epsilon\) is called the leakage [6]. Inequality
(1) implies that the pattern mining algorithm outputs are $\epsilon$-indistinguishable probabilistically when there is only one different entry in the database, $\epsilon$ has to be small, since \( \log(1 + \epsilon) \approx \epsilon \) for small $\epsilon$. In other words, a high degree of differential privacy is quantified by a small leakage.

$\epsilon$-differential privacy can be achieved by the addition of random noise whose magnitude is chosen as a function on the largest change a single participant could have on the output to the query function, called the sensitivity of the function.

**Definition 3:** For $f : D \rightarrow R^d$ such that $D$ is the database and $R^d$ is a $d$-dimensional vector space, the sensitivity of $f$ is

$$\Delta f = \max_{D_1, D_2} ||f(D_1) - f(D_2)||_1$$

for all $D_1, D_2$ differing in at most one element.

The most common mechanism to handle differential privacy is by Laplace noise perturbation on the outputs [6] described by Theorem 1 below.

**Theorem 1:** Let $K_f$ be the privacy mechanism for a query function $f : D \rightarrow R^d$ and adds noise with a scaled symmetric exponential distribution with variance $\sigma^2$ described by the density function

$$Pr[K_f(X) = a] \propto \exp \left( -\frac{||f(X) - a||_1}{\sigma} \right)$$

to each component of $f(X)$ with $X \in D$. The density function defines the Laplace distribution, Lap($\sigma$). The mechanism $K_f$ gives \( \left( \frac{\Delta f}{\sigma} \right) \)-differential privacy.

The proof of Theorem 1 shows that the application of Laplace noise satisfies Inequality (1) [28]. To achieve $\epsilon$-differential privacy, one perturbs the query output $f(X)$ so that the privacy preserved output

$$f'(X) = f(X) + N$$

where $N \sim$ Lap$(\sigma)$ with $\sigma = \Delta f / \epsilon$.

One important result for the differential privacy concept is that the application of a sequence of differential privacy mechanisms ensures differential privacy for a task, even when a later computation makes use of results from earlier computations [24].

**Theorem 2 (Sequential Composition):** Let $M_i$ be a mechanism that provides $\epsilon_i$-differential privacy for $i = 1, \ldots, n$. The sequence of $M_i$ on an input domain $D$ provides \( \sum_{i=1}^n \epsilon_i \)-differential privacy.

Another important result is that if a sequence of differential privacy mechanisms is applied to a set of disjoint subsets of the input domain, one achieves a better overall privacy guarantee which is the worst privacy guarantee among all the disjoint subsets.

**Theorem 3 (Parallel Composition):** [24] Let $M_i$ be a mechanism that provides $\epsilon_i$-differential privacy for $i = 1, \ldots, n$ on arbitrary disjoint subsets $D_i$ of the input domain $D$. The sequence of $M_i$ on the input domain $D$ provides \( \max \epsilon_i \)-differential privacy.

A “relaxed” definition for differential privacy used in [29] allowing one to claim the same privacy level as Definition 1 when there is a small amount of privacy loss due to a slight variation in the output distribution for the privacy mechanism $K$ is as follows.

**Definition 4:** [29] A randomized function $K$ is $(\epsilon, \delta)$-differentially private if for all datasets $D_1$ and $D_2$ differing on at most one element, and all $S \subset \text{Range}(K)$

$$Pr[K(D_1) \in S] \leq \exp(\epsilon) \times Pr[K(D_2) \in S] + \delta.$$ 

One notes that if $\delta = 0$, $(\epsilon, 0)$-differential privacy is $\epsilon$-differential privacy. This is the differential privacy definition we use for the rest of the paper. Moreover, one notes that this definition has no relation to the $(\epsilon, \delta)$-probabilistic differential privacy [30] such that $\delta$ is used to ensure that the disclosure set has low probability.

The above composition theorems are also valid for $(\epsilon, \delta)$-differential privacy mechanism. The sensitivity $\Delta f$ used to specify the $\sigma$ parameter for the Laplace distribution (Theorem 1) is a “global sensitivity” [31] as the sensitivity depends on all possible pairs of $D_1$ and $D_2$ with one element difference. In other words, the calculation of $\Delta f$ is not based on a particular database instance. Instead, one has to derive a database independent sensitivity for a query function $f$. Due to this database independent nature of sensitivity, the privacy mechanism can perform very badly in practice [32]. In particular, if the noise perturbation is too much due to large sensitivity, the output responses from the algorithm becomes meaningless.

Next, we describe and discuss challenges related to applying differential privacy mechanism on moving objects database mining tasks and describe the solution we proposed in [33]. In particular, we use the frequent location pattern mining task as an example for illustration.

**V. Privacy Solution**

In this paper, we focus on the privacy issue related to frequency or support count, $\sigma(X)$ for the three mining tasks described in Section III.

A practical problem for the Laplace noise perturbation privacy mechanism (3) is the magnitude of the database independent global sensitivity, $\Delta f$. For the frequent location pattern mining task, the criterion to identify a frequent location is the number of stay points for a particular location in the database. However, one cannot derive a reasonably useful database independent $\Delta f$ for the count query from a stay point database. The key issue is that the database independent global sensitivity $\Delta f$ for the count query does not equal one when an individual is added or removed from the database. The removal of an individual’s record from the database can result in no stay point removed from a frequent location pattern or the frequent location pattern no longer exists (as too many stay points removed). Similarly, the addition of an individual’s record from the database can result in no stay point added to a frequent location pattern or a new frequent location pattern discovered (as many stay points are added to a new location). Hence, without any additional assumptions on the pattern
mining task, there is no other possible $\triangle f$ except the total number of stay points computed from the trajectories. This $\triangle f$ is practically useless as the global sensitivity dependent noise (3) will create meaningless count query output.

To partially overcome some of the situations mentioned above, we include two assumptions into our frequent location pattern mining algorithm as follows.

A1. There should be at least $m$ (say $m'' > 1$) individuals in a frequent location pattern; and

A2. There should be at least $r$ stay points for each individual for at least $m'(\leq m'')$ individuals in a frequent location pattern.

From assumptions A1 and A2, one notes that there need to be at least $m$ individuals but not all have at least $r$ stay points. The count query $f$ considers counts from all $m''$ individuals even when the number of stay points for an individual is less than $r$. However, the two assumptions still do not allow us to compute or define a database independent global sensitivity $\triangle f$. Hence, the motivation for our solution is to utilize database dependent local sensitivity to parametrize the noise used for output perturbation in the frequent location pattern mining task.

**Definition 5:** [31] For $f: D \to R^d$ and $x \in D_1$, the local sensitivity of $f$ and $x \in D$ is

$$\triangle f_{LS}(x) = \max_{y \in D} ||f(x) - f(y)||_1$$

for all $y \in D$ differing in at most one element from $x \in D$. According to [31], the noise function has to be insensitive in the database space. In other words, a small difference between two databases cannot induce a spike or sharp dip in the noise added. [31] introduced the concepts of $\beta$-smooth sensitivity and $\beta$-smooth upper bound for local sensitivity.

**Definition 6:** [31] For $\beta > 0$, the $\beta$-smooth sensitivity of $f$ for $x \in D$ is

$$\triangle f_{\beta}(x) = \max_{y \in D}(\triangle f_{LS}(y) \cdot e^{-\beta d(x,y)})$$

such that $d(x,y)$ is the number of elements that $x$ and $y$ differs. One notes that it is difficult to compute $\triangle f_{\beta}(x)$ as one has to find local sensitivity for all databases in $D$.

**Definition 7:** [31] For $\beta > 0$, a function $S: D \to R^+$ is a $\beta$-smooth upper bound on $\triangle f_{LS}$, local sensitivity of $f$, if it satisfies the following requirements:

$$\forall x \in D : \quad S(x) \geq \triangle f_{LS}(x)$$

$$\forall x, y \in D : \quad S(x) \leq e^{\beta} S(y)$$

such that $x$ and $y$ differing on one element.

The noise added to the output is weighted by the smooth upper bound function $S$ on the local sensitivity for all databases in $D$ such that $\ln S(\cdot)$ has a global sensitivity bounded by $\beta$. Note that $\triangle f_{\beta}$ is a $\beta$-smooth upper bound on $\triangle f_{LS}$ and the global sensitivity $\triangle f$ is, in general, a relax/conservative $\beta$-smooth upper bound. Moreover, the noise distribution $N$ on $R^d$ (e.g., Laplace distribution) should not change much under translation (T) and scaling (S) [31], i.e., when $N_1, N_2 \sim N$,

T: $P(N_1 \in U) \leq e^\frac{\epsilon}{\sigma} P(N_2 \in U + T) + \frac{\delta}{2}$

S: $P(N_1 \in U) \leq e^\frac{\epsilon}{\sigma} P(N_2 \in e^\beta U + T) + \frac{\delta}{2}$

for all $||T|| \leq \alpha$, and $\lambda \leq \beta = \beta(\epsilon, \delta)$, all subsets $U \subseteq R^d$, $\epsilon > 0$, and $\delta > 0$. A noise distribution $N$ is $(\alpha, \beta)$-admissible if it satisfies the above two conditions.

Now, we can achieve $(\epsilon, \delta)$-differential privacy using the $(\alpha, \beta)$-admissible Laplace noise by using the privacy mechanism

$$K(x) = f(x) + \frac{S}{\alpha} \cdot N$$

such that $\triangle f_{\beta}$ is the $\beta$-smooth upper bound on $\triangle f_{LS}$, local sensitivity of $f$ and the noise random variable $N \sim Lap(1) = \frac{1}{\frac{1}{2}} \cdot e^{-|x|}$, $\alpha = \frac{1}{2}$, and $\beta = \frac{1}{2} \ln(\frac{1}{\beta})$ with $\delta > 0$.

The above privacy mechanism follows from Lemma 2.5 in [31] that provides the condition for $(\epsilon, \delta)$-indistinguishable for $\triangle f_{LS}$ and local sensitivity of $f$, together with Example 3 in [31] describing a $(\alpha, \beta)$-admissible Laplace distribution with location parameter $\mu = 0$ and scaling parameter $\sigma = 1$ that satisfies Definition 4 with $\delta > 0$. The differential privacy level $\epsilon$ and the local sensitivity $\triangle f_{LS}$ are no longer used to parametrize the Laplace distribution (3). Instead, they are used to weigh the noise generated from a standard Laplace distribution with mean zero and variance one.

To use this privacy mechanism for the frequent location pattern task, one need to compute either the $\beta$-smooth sensitivity or the $\beta$-smooth upper bound. We compute the later for the count query $f$ as it is achievable in practice by adding constraints to the frequent location pattern mining algorithm. We add the following assumption.

A3. If an individual has more than $r_{\text{max}}$ stay points at a frequent location $i$, the number of stay points in $i$ for the individual is set to $r_{\text{max}}$.

The local sensitivity $\triangle f_{LS}$ for the count query following Definition 5 and also satisfying assumption A3 can be computed as follows.

$$\triangle f_{LS} = \max_{u \in U, i \in I_u} |o_{iu} - n_{iu}|$$

such that $I_u U$ is the set of frequent locations when an individual $u$ and all its corresponding stay points are removed or added to a database, and $o_{iu}$ and $n_{iu}$ are the number of stay points in such a frequent location $i$ before and after the removal or addition of stay points for an individual $u$.

To satisfy both conditions in Definition 7 for $\triangle f_{LS}$, we have the constant function

$$S(x) = r_{\text{max}}$$

as the $\beta$-smooth upper bound on $\triangle f_{LS}$.

Substituting (9) into (6), one obtains a $(\epsilon, \delta)$-differentially private count query for the frequent location pattern mining task using (6). More theoretical and empirical results are found in [33].
VI. DISCUSSIONS

Our proposed privacy mechanism (6) discussed in the previous section has the advantage of a smaller perturbation magnitude than the conventional Laplace noise privacy mechanism (3) and at the same time, maintaining the privacy level. It can be applied to the other two moving objects data mining tasks: sequential pattern mining and association rule generation. For these tasks, the output patterns/rules remain unchanged. The differential privacy mechanism is also applied to the support count. In such a way, the output patterns/rules remain unchanged but an individual’s privacy is preserved. In other words, one cannot tell whether a frequent pattern, sequential pattern or association rule is related to an individual.

For anomaly or outlier detection in moving objects database, the task objective is to detect trajectories that are abnormal in the sense that they deviate from normal behavior in terms of frequent/regular locations or travel sequences. One notes that assumption A1 and A2 (see Section V) make the differential privacy solution impossible to satisfy the task objective. Moreover, the objective of differential privacy (and any privacy goal, in general) also contradicts the objective of anomaly or outlier detection. One open question is whether it is possible to handle privacy issues in terrorism related analytics and crime forecasting that utilize anomaly or outlier detection.

VII. CONCLUSIONS

In this paper, we describe and discuss how one can ensure an individual’s privacy with respect to his location and spatiotemporal behavioral patterns via differential privacy mechanism; in particular, we demonstrate our recently proposed privacy preserving approach on the frequent location pattern mining task. We also discuss the extension of its use to sequential pattern mining and association rules generation, and the dilemma of differential privacy goal for anomaly or outlier detection in moving objects data mining.

REFERENCES


