SLICER: A Slicing-Based K-Anonymous Privacy Preserving Scheme for Participatory Sensing

Fudong Qiu, Fan Wu, Guihai Chen
Shanghai Key Laboratory of Scalable Computing and Systems
Department of Computer Science and Engineering
Shanghai Jiao Tong University, China
Email: fdqiu@sjtu.edu.cn, {fwu,gchen}@cs.sjtu.edu.cn

***Abstract*—With the popularity of mobile wireless devices with various kinds of sensing abilities, a new service paradigm named Participatory Sensing has emerged to provide users with brand new life experience. However, the wide application of participatory sensing has its own challenges, among which privacy preservation and multimedia data participatory sensing are two critical problems. Unfortunately, none of the existing works has fully solved the problem of privacy preserving participatory sensing with multimedia data. In this paper, we propose SLICER, which is the first k-anonymous privacy preserving scheme for participatory sensing with multimedia data. SLICER integrates a data coding technique and message exchanging strategies, to achieve strong protection of participants' privacy, while maintaining high data accuracy. In addition, two slice transferring strategies are well designed for slice transfer to minimize the total transfer cost. Finally, we have implemented SLICER and evaluated its performance using publicly released taxi traces. Our evaluation results show that SLICER achieves high data accuracy, with low computation and communication overhead.**

I. Introduction

The wide application of mobile communication equipments and the fast advance of sensing technologies have led to the wide availability of privately-held, low-cost, advanced-processing, and big-storage mobile wireless devices, that are equipped with a number of embedded sensors (*e.g.*, microphone, camera, accelerometer, gyroscope, and GPS). On one hand, modern wireless communication technologies (*e.g.*, 2G/3G/4G, Wi-Fi, and Bluetooth) make the communication between mobile devices and infrastructure, as well as between mobile devices themselves, convenient and fast. On the other hand, the mobile devices, especially smart phones, are no longer a tool only for communication, but "computers" with multifunction.

Participatory Sensing [1] emerged as a new service paradigm using human-carried mobile devices, such as smart phones, for distributed data collection, analysis, and sharing. With an estimated number of 5.9 billion mobile phones worldwide [2], participatory sensing may provide an unprecedented spatial coverage, with low or no deployment cost. Compared with traditional decentralized data collection methods (*e.g.*, Wireless Sensor Networks), participatory sensing demonstrates several outstanding advantages, including larger coverage, lower cost, more sufficient energy supply, and more flexible interactive capability. Attracted by the practical and commercial value of participatory sensing, many participatory sensing applications have appeared. For instance, GreenGPS [3] provides the most fuel-efficient routes to drivers; PEIR [4] presents a personal environmental impact report for every individual; and Ikarus [5] uses sensor data collected during cross-country flights via participatory sensing applications to study thermal effects in the atmosphere, and PoolView [6] gives a privacy preserving architecture for stream data collection.

However, the application of participatory sensing has a number of challenges. One of the major challenges is on privacy preservation. Sensing record sent to the service provider, is usually attached with spatio-temporal tags indicating the location and time of the data collected. However, a corrupt service provider may infer private information of the participants, such as identity, home and office addresses, traveling paths, as well as participants' habits and lifestyles, from the sensing records. In turn, many users are reluctant to contribute any sensing record if proper privacy preservation scheme is not applied. Without sufficient number of participants, participatory sensing applications cannot guarantee their quality of services at the expected level. Therefore, designing privacy preserving schemes for participatory sensing is highly important. Another major challenge is on the variety of sensing data. Most of existing applications of participatory sensing only collect small pieces of sensing data (*e.g.*, temperature, velocity, and geographic location). However, more and more newly emerged applications rely on collecting information of surrounding environment in the format of multimedia (*e.g.*, digital image and video) [7], which result in much higher volume of sensing data. Simply applying existing privacy preserving schemes to participatory sensing with multimedia data is not satisfactory since existing schemes either induce unacceptable amount of communication cost, or degrade the utility/accuracy of the data badly, in case of multimedia sensing.

In this paper, we present SLICER, which is a k-anonymous privacy preserving scheme, working on application layer, for participatory sensing with multimedia data. Intuitively, k-

This work was supported in part by the State Key Development Program for Basic Research of China (Grant No. 2012CB316201), in part by China NSF grant 61272443, 61133006, 61073152, and in part by Shanghai Science and Technology fund 12PJ1404900 and 12ZR1414900. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.

F. Wu is the corresponding author.

978-0-7695-5104-3/13 $26.00 © 2013 IEEE
DOI 10.1109/MASS.2013.33

anonymity means that the service provider cannot identify the contributor of each sensing record from a group of at least k participants. SLICER integrates a data coding technique and message exchanging strategies, to achieve strong protection of participants' privacy, while maintaining high data accuracy and inducing low communication and computation overhead.

The contributions of this work are listed as follows:

- We propose SLICER for participatory sensing with multimedia data, to achieve both k-anonymous privacy preserving and high data accuracy, with low communication and computation overhead.
- We design an erasure coding based sensing record slicing scheme to encode each sensing record into a number of data slices, each of which can be delivered to the service provider through the other participants or the record's generator herself. When a proper data slice exchanging strategy is applied, the contributor of each particular sensing record is hidden in a group of at least k participants.
- We propose two kinds of strategies for slice transfer. The first and straight forward strategy is to transfer a slice upon meeting another participant. The later delivers the slice to the service provider. The second one is an approximately efficient strategy to transfer the slices to a set of participants that might be met within a required period of time, minimizing the total cost while guaranteeing that the sensing record can be delivered to the service provider with high probability.
- We have implemented SLICER and evaluated its performance using publicly released real traces of taxis [8]. Evaluation results show SLICER achieves high data accuracy, with low computation and communication overhead.

The rest of this paper is organized as follows. In section II, we introduce technical preliminaries, including system model, privacy model, and design objectives. In section III, we present our design of SLICER in details. In section IV, we present evaluation results. In section V, we discuss the related works. Finally, we conclude our paper and point out future work directions in section VI.

II. Technical Preliminaries

In this section, we present the system model, privacy models, as well as objectives of our design.

A. System Model

We consider a cloud-based participatory sensing and service framework as shown in Fig. 1, in which there is a service provider and a number of mobile nodes/participants.

The service provider aggregates, classifies, analyzes, and stores sensing records reported from the participants, and provides query services based on the records. A mobile node/participant is a user carrying a portable and wireless-enabled device (*e.g.*, smart phone, tablet, and laptop). In this paper, we use mobile node and participant interchangeably. Participants can use their sensing devices to collect various kinds of environmental information, such as geographical location, temperature, electromagnetic signal, digital image, video, and so on. In contrast to most of the existing works, which focus on short sensor readings, we consider a participatory sensing system that adapts to multimedia information, such as digital image and video. The participants can directly report sensing records through pre-existing communication infrastructure, including GSM, 3G/4G, and Wi-Fi, or indirectly report the records with the help of the other participants.

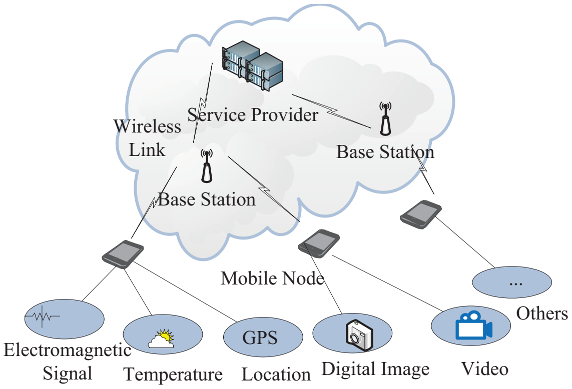

Fig. 1. The Architecture of Cloud-Based Participatory Sensing.

B. Privacy Model

Although participatory sensing provides a new service paradigm, its functionality relies on the contribution of participants. Existing works [1], [9]–[14] show that contributed information may be misused to reveal the participants' privacy. Most users are not willing to join participatory sensing applications, unless their sensitive information is well protected.

In this paper, we consider the problem of privacy preserving in a semi-honest model, in which the adversary correctly follows the protocol specification, but attempts to learn additional information by analyzing the transcript of messages received during the execution [9], [15]–[18]. We classify the attacks in the semi-honest model into two categories: external attack and internal attack. The external attack aims to obtain private information of participants by overhearing the message passing through the wireless communication network. Such attack can be prevented by end-to-end cryptographic schemes. Different from the external attack, designing a scheme to prevent the internal attack is much more challenging. The internal attack may come from two different kinds of entities, including the service provider and the participants.

- Service provider's attack: The service provider has full access to the sensing records reported by the participants. It might infer considerable amount of sensitive information about the participants (*e.g.*, home address, traveling path, and lifestyle), if a proper privacy-preserving scheme is not provided. For instance, the sensor readings collected by a user who drives from home to work might reveal the participant's traveling path as well as her home address. In this work, we focus on protecting users' location/path privacy against the service provider, while assuming that the service provider does not have other background or correlated information about participants. It is also important to consider the privacy protection of the content of multimedia data. However, it is out of the scope of this work. For interested readers, please refer to [19]–[21] for privacy processing techniques.

- Participants' attack: Participants may receive some sensing records, when they serve as relays for other participants (*e.g.*, in [22]). Semi-honest participants might position themselves to some critical locations in order to collect sensitive information by pretending to be relays. In this work, we assume that the participants do not collude with the service provider.

C. Design Objectives

In this paper, we consider a set $N = \{a_1, a_2, \ldots, a_n\}$ of participants. Each participant $a_i \in N$ would like to contribute her sensing records $R_i = \{< t_1, l_1, d_1 >, < t_2, l_2, d_2 >, \ldots\}$ to the service provider, only when her privacy is properly protected. The triple $< t, l, d >$ denotes a sensing record including *timestamp*, *location info*, and *data info*.

The design of privacy preserving scheme should prevent both the external and the internal attacks. Specifically, first, the design needs to prevent external eavesdroppers from obtaining any meaningful information. Second, the design needs to prevent service provider from recognizing the identity of the participant who contributes a particular sensing record, and to prevent the participants from knowing the content of the relayed sensing record. Especially, we require the privacy protection scheme be k-anonymous [23] against the service provider. Here, k-anonymity is reached when the service provider can only identify a particular participant that contributes a sensing record with probability no more than $1/k$.

Definition 1 (K-Anonymous Participatory Sensing): A privacy preserving participatory sensing scheme satisfies k-anonymity against the service provider, if for any sensing record reported to the service provider, the service provider cannot distinguish the the generator of the record from a group of at least k participants.

Besides the objective on privacy preservation, the design should also satisfy the following requirements:

- The design should maintain high accuracy of the sensor readings.
- The design should be tolerant of packet/message loss.
- The design can only induce low computation and communication overhead.

III. Slicing-based Privacy Preserving Scheme

In this section, we present the design of our slicing-based k-anonymous privacy preserving scheme — SLICER. We first outline the general idea of SLICER, and then explain the details of each component. Finally, we analyze the privacy preservation properties of SLICER.

A. Design Rationale

The main idea of SLICER is to hide the generator of each sensing record among a group of at least k participants, through which all parts of the sensing record are reported to the service provider. Thus, the service provider cannot identify the generator of the sensing record from at least k participants. We illustrate the designing challenges and our idea in this section.

(1) Sensing Record Slicing

If we simply transfer the (encrypted) sensing record to k participants, then the communication overhead is k times the size of the sensing record, which is unacceptable especially when the sensing record contains multimedia data. Therefore, we incorporate erasure coding to encode each sensing record into a number of small slices. Then each of the slices can be transferred to a participant, and the later reports the slice to the service provider. Once the service provider receives enough number of slices, not necessarily all the slices, it can decode the original sensing record. The usage of erasure coding has two advantages. One is to greatly reduce the communication overhead needed to transfer the sensing record (slices in this paper) to other participants. The other is to increase the reliability of the system, when the slices may be lost due to various reasons.

(2) Transfer Strategy

Since the slices need to be transferred to a set of participants, carefully selecting the participants to transfer to may affect the performance of the scheme. The straight forward strategy is to transfer a slice whenever another participant is met. However, when the participants in the system have different capabilities, the straight forward way may not be the best strategy. In this paper, we consider the case, in which the participants have different cost to deliver a slice. The cost difference can be resulted from the wireless communication fee, available bandwidth, battery power, and so on. We also propose two sub-optimal slice transfer strategies to minimize the total cost for delivering the slices.

Fig. 2 shows the general work flow of SLICER. A sensing record contains the sensor reading and spatio-temporal information. Then, SLICER encodes the sensing record using erasure coding, encrypts the coded blocks, and attaches an unique tag, to generate slices. Next, SLICER selectively transfers the slices to the participants met, following one of its transfer strategies. The slices are delivered to the service provider through different participants. Finally, the service provide decrypts the slice and reconstructs the original sensing record, when enough number of slices are received.

In the following subsections, we present the design details of SLICER's major components, including slicing, transferring, and reconstructing.

B. Slicing

Algorithm 1 shows the pseudo-code of our sensing record slicing algorithm. Given a sensing record $< t, l, d >$ from participant $a_i \in N$, we encode it into a number of slices, each of which will be delivered to the service provider through

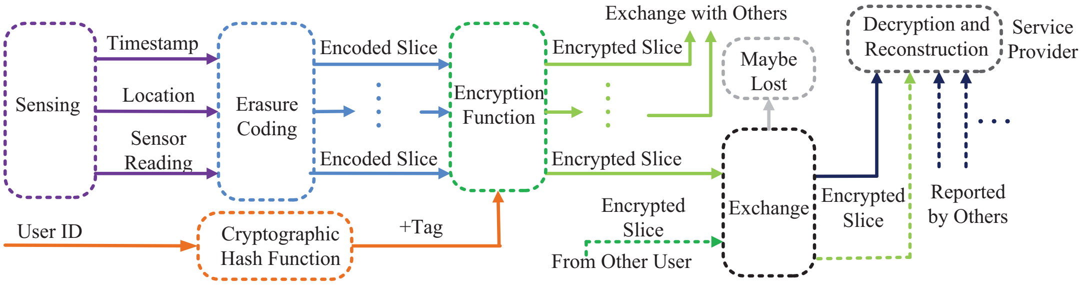

Fig. 2. Work Flow of SLICER.

```
Algorithm 1 Sensing Record Slicing Algorithm
Input: A sensing record $< t, l, d >$ from participant $a_i \in N$,
    and coding rate $k/m$.
Output: Encrypted slices $\{r'_{ij} | 1 \leq j \leq m\}$.
$\{r_{ij} | 1 \leq j \leq m\} \leftarrow EC(< t, l, d >)$;
$nonce \leftarrow random()$;
$tag = H(i, nonce)$;
for all $j = 1$ to $m$ do
    $r'_{ij} = ENCRYPT(r_{ij} || tag, KEY_{pub})$;
end for
return $\{r'_{ij} | 1 \leq j \leq m\}$;
```

different participants. We encode the record $< t, l, d >$ using erasure coding (*e.g.*, Reed-Solomon [24] and Tornado [25]). Basically, erasure coding divides a record into k slices, and encodes them into m slices, where $m > k$. The original record can be reconstructed from any k out of m encoded slices. The ratio k/m is the coding rate. Here the combined size of any k slices is approximately equal to the size of the original record. Intuitively, if the service provider decodes the record from k slices reported by k different participants, the real generator of the record is hidden in a group of k participants, which provides a privacy guarantee of k-anonymity. Furthermore, SLICER inherits the property of loss tolerance from erasure coding to achieve high record reconstruction ratio with relatively lower communication overhead. We denote the encoded slices by $\{r_{ij} | 1 \leq j \leq m\}$:

$$\{r_{ij} | 1 \leq j \leq m\} = EC(< t, l, d >),$$

where $EC(\cdot)$ is one of the erasure coding algorithms.

Since the service provider may receive a large number of encoded slices originating from various participants' sensing records, we have to tag the slices to clearly indicate which slices belongs to the same record. Since directly tagging a slice with its generator's ID and a sequence number will reveal the identity privacy of the generator, we adopt a cryptographic hash function (*e.g.*, SHA-1 [26]) to create the tag:

$$tag = H(i, nonce),$$

where $H(\cdot, \cdot)$ is a cryptographic hash function and $nonce$ is an arbitrary number.

To prevent the content of encoded slices being revealed to external attacker and neighboring participants, we encrypt the encoded slices and the tag using the public key KEY_{pub} of the service provider and get the encrypted slices:

$$r'_{ij} = ENCRYPT(r_{ij} || tag, KEY_{pub}), 1 \leq j \leq m,$$

where $ENCRYPT(\cdot, \cdot)$ is an asymmetric encryption function, and $||$ is string concatenation operation.

C. Transferring

To prevent the service provider from recognizing participants' identities with the collected sensing records, not all slices of a sensing record can be directly sent to the service provider by the generator. To guarantee k-anonymity, at least $k - 1$ slices need to be delivered by participants other than the generator. We note that although all the slices can be transferred to and delivered by participants other than the generator, SLICER requires the generator to report (at least) one slice to the service provider by herself, in order to guarantee the integrity of the sensing record.

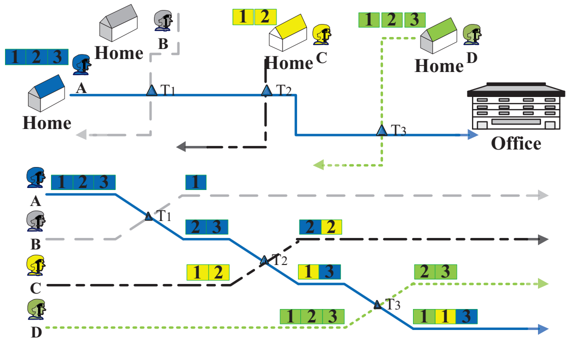

Fig. 3. An Example of Transfer on Meet Up

In this paper, we consider two kinds of slice transferring strategies: *transfer on meet up(TMU)* and *minimal cost transfer(MCT)*.

1) Transfer on Meet Up (TMU): This is the straight forward way to spread the slices. One slice of each sensing record is transferred, when the generator meets another participant. Later, the participants, including the generator, report the slices to the service provider.

Fig. 3 shows a toy example of applying the strategy of TMU. Assume that there is a participant A who is going to office from her home. She meets other participants B, C, and D in sequence on her way to the office. The upper part of Fig. 3 shows the path that A travels, and the lower part shows the slices each of the users hold with advance of time. Assume that A, B, C, and D initially has 3, 0, 2 and 3 slices of their own, respectively, and meetings occurs at T_1, T_2, and T_3, at where a participant transfer one slice to the one met. For example, at T_1, A transfers one slice to B. After that, A has 2 slices left, and B holds 1 slice from A. Finally, after three meetings, A has 1 own slice and 2 slices from C and D, B has 1 slice from A, C has 1 own slices and 1 from A, and D has 2 own slices.

2) Minimal Cost Transfer (MCT): Here, we present our solution for the problem of Minimal Cost Transfer (MCT).

Each sensing record has an expiration time, before which the record has to be delivered to the service provider. We assume that each participant $a_i \in N$ knows a set $N(a_i) \subset N$ of participants that might be met before the expiration of the sensing record. For each participant $a_j \in N(a_i)$, let $p(a_j)$ and $c(a_j)$ be the meeting probability before the expiration time and the cost of the participant a_j for delivering a slice. As we mentioned before, the cost can be resulted from the wireless communication fee, available bandwidth, battery power, and so on. We assume that there is a mobility prediction module (*e.g.*, [27], [28]) to provide the prediction of $N(a_i)$, $(p(a_j))_{a_j \in N(a_i)}$, and $(c(a_j))_{a_j \in N(a_i)}$, based on historical event logs.

The objective of MCT is to pick a subset of participants $F \subseteq N(a_i)$ as forwarders of the slices to minimize the cost for delivering the slices, satisfying the following requirement:

- Requirement: It is expected to meet at least $m-1$ participants from the forwarder set F, namely MCT-EXP problem;

Next, we will present our approach to solve the above MCT-EXP problem.

Solution to MCT-EXP Problem

We first consider the MCT-EXP problem (*i.e.*, MCT problem with requirement 1), which can be formulated as a binary program with an objective of minimizing the expected delivery cost of the slices, as follows:

Objective:

$$Minimize \qquad \sum_{a_j \in N(a_i)} (c(a_j)p(a_j)x_j)$$

Subject to:

$$\sum_{a_j \in N(a_i)} (p(a_j)x_j) \geq m-1, \tag{1}$$

$$x_j \in \{0,1\}, \qquad \forall a_j \in N(a_i) \tag{2}$$

Here, constraint (1) guarantees that participant a_i is expected to meet at least $m-1$ participants in the selected forwarder set $F = \{a_j \in N(a_i) | x_j = 1\}$. Constraint (2) indicates the possible values of x_j. If a_j is selected to be a candidate for delivering a slice, then $x_j = 1$; otherwise, $x_j = 0$.

We note that the above formulation of MCT-EXP Problem can be reduced to the 0-1 Knapsack Problem [29] with an objective of maximizing the expected cost of the complimentary of the forwarder set, as follows:

Objective:

$$Maximize \qquad \sum_{a_j \in N(a_i)} (c(a_j)p(a_j)(1-x_j))$$

Subject to:

$$\sum_{a_j \in N(a_i)} (p(a_j)(1-x_j)) \leq \sum_{a_j \in N(a_i)} p(a_j) - (m-1), \tag{3}$$

$$x_j \in \{0,1\}, \qquad \forall a_j \in N(a_i) \tag{4}$$

In the reduced 0-1 Knapsack Problem, $p(a_j)$ and $c(a_j)p(a_j)$ are the weight and value of the jth item, respectively, while the capacity of the knapsack is $\sum_{a_j \in N(a_i)} p(a_j) - (m-1)$. Here, constraint (3) guarantees that the sum of the weights must be less than the knapsack's capacity. Constraint (4) is exactly the same as constraint (2). Consequently, we can have a Fully Polynomial Time Approximation Scheme (FPTAS) [29], which runs in polynomial time and is correct within $1-\epsilon$ percent of the optimal solution, to solve the MCT-EXP problem. Due to limitations of space, we refer the reader to [29] for the detailed solution.

The above strategy can return a feasible result if there are sufficient number of meeting opportunities with other participants. However, we note that it is possible that a sensing record generator cannot meet enough participants to transfer each of the encoded slices from a record to a different participant. In this case, we use the prediction model based on the history to estimate the number of encounters beforehand. For participants who do not have sufficient slice transfer opportunities, we allow them to transfer more than one slice during each meeting. Suppose h slices are transferred each time, then the record generator is hidden in $\lceil k/h \rceil$ participants.

D. Reconstructing

After receiving at least k slices encoded from the same sensing record, the service provider can reconstruct the original sensing record. Besides maintaining a database storing the sensing records, the service provider also keeps a table T caching slices that have not been decoded.

Algorithm 2 shows the pseudo-code of our sensing record reconstructing algorithm. Upon receiving a reported slice s, the service provider decrypts the slice using her private key KEY_{priv} to get the encoded slice s' and a tag that uniquely identifies the record it is encoded from:

$$(s', tag) = DECRYPT(s, KEY_{priv}),$$

where $DECRYPT(\cdot,\cdot)$ is an asymmetric decryption function.

The service provider adds the encoded slice s' into the cache table T with index tag, and then check whether there are

Algorithm 2 Sensing Record Reconstructing Algorithm

```
Input: Cache table $T$.
Output: Each original sensing record $< t, l, d >$.
while $TRUE$ do
    Receive slice $s$;
    $(s', tag) \leftarrow DECRYPT(s, KEY_{priv})$;
    Add $(s', tag)$ into $T$;
    if $|\{\bar{s}| < \bar{s}, \bar{t} > \in T \wedge \bar{t} = tag\}| \geq k$ then
        $< t, l, d > \leftarrow EC^{-1}(\{\bar{s}| < \bar{s}, \bar{t} > \in T \wedge \bar{t} = tag\})$;
        Remove $\{\bar{s}| < \bar{s}, \bar{t} > \in T \wedge \bar{t} = tag\}$ from $T$;
        Store sensing record $< t, l, d >$;
    end if
end while
```

k encoded slices with the same tag. If so, she extracts the k encoded slices with the same tag, and then decodes the original sensing record:

$$< t, l, d >= EC^{-1}(\{\bar{s}| < \bar{s}, \bar{t} > \in T \wedge \bar{t} = tag\}),$$

where $EC^{-1}(\cdot)$ is the decoding function corresponding to $EC(\cdot)$.

E. Analysis

In this section, we show that SLICER can provide strong privacy protection against the external and internal attacks.

1) Protection Against External Attacks: The external attacker eavesdrops messages passed in the participatory sensing system, in order to collect sensitive information about particular participants. In SLICER, we employ an end-to-end cryptographic encryption scheme, such that the external attacker cannot decrypt the slices transferred among participants, as well as that reported to the service provider. Although the external attacker may extract some information from the eavesdropped packets to uniquely identify the participant, she cannot get the content of the sensing record. Therefore, SLICER provides privacy protection against the external attacks.

2) Protection Against Internal Attacks: The internal attack may come from both the participants and the service provider. We distinguish two cases:

Protection against participants' attack

Each participant may receive some slices, when she is selected as a slice deliver for participants met. Similar with the external attacker, the participant cannot decrypt the slice for delivering.

Protection against service provider's attack

Since the service provider has full access to the sensing records contributed by the participants, she can easily infer private information about the participants, if proper privacy-preserving scheme is not provided. However, SLICER can achieve the k-anonymity and protect participants' privacy information against the service provider. Therefore, we can draw the following theorem.

Theorem 1: SLICER achieves k-anonymity, when there are k participants who deliver slices to the service provider.

Proof: In SLICER, we isolate the participants' identity and the sensing records, by encoding each sensing record into m slices and letting at least k different slices be delivered to the service provider through different participants. To achieve this, we designed two different algorithms (TMC and MCT-EXP) in section III-C according to different situations to select at least m participants (including the generator itself) as forwarders to transfer m slices to the service provider. Then, the original sensing record can be decoded by the service provider if and only if receiving at least k different slices. Therefore, the identity of the record generator is hidden among a group of at least k participants. ■

We note that SLICER's privacy guarantee degrades to $\lceil k/h \rceil$-anonymity, when a sensing record generator cannot meet enough participants to transfer slices and thus has to transfer h slices during each meeting. Further, if the sensing record generator is completely isolated and cannot meet any other participant (*i.e.*, $h = k$), SLICER cannot preserve the privacy on linkage between identity and location. In this case, an alternative privacy preserving scheme (*e.g.*, [10], [11]) can be applied.

IV. Evaluation

We have implemented SLICER and evaluated its performance on real world taxi traces. In this section, we present our evaluation results.

A. Setup and Metrics

Our evaluation is based on the realistic GPS mobility traces of 500 taxi cabs over 30 days in San Francisco, which were collected by Cabspotting [8] and can be accessed from the CRAWDAD [30]. In this real world deployment, each cab is outfitted with a GPS tracking device that is used by dispatchers to efficiently reach customers. Each cab sends a location-update (timestamp, identifier, geo-coordinates) to a central server in a period varied from 30 to 60 seconds, which forms the mobility traces we used in this paper. We extend this scenario to a participatory sensing situation by assuming that the cabs are participants equipped with mobile devices.

We consider a mobile infrastructure with the whole 500 participants. We set that every participant generates one record per day, and the valid period of the record is 24 hours. The loss possibility of the slices varies from 0.2 to 0.4. The value of k is set to 10.

We evaluate the performance of SLICER using four metrics:

- *Reconstruction Ratio*: the percentage of sensing records successfully reconstructed by the service provider. This reflects the loss tolerance of SLICER.
- *Communication Overhead*: the total amount of data transmitted to guarantee required reconstruction ratio.
- *Computation Overhead*: the time consumed to process a sensing record.
- *Total Transfer Cost*: the sum of the cost for delivering a sensing record (*i.e.*, $m - 1$ slices) to the service provider.

B. Evaluation Results

We compare the performance of SLICER implemented with two transfer strategies proposed in section III (namely *TMU* and *MCT-EXP*), with an existing privacy preserving schemes for participatory sensing, namely *Simple Exchanging* [22], in which the sensing records are transferred among participants as a whole without slicing/coding. We should note that we did not compare with [10]–[12], because the setup of these are significantly different with ours.

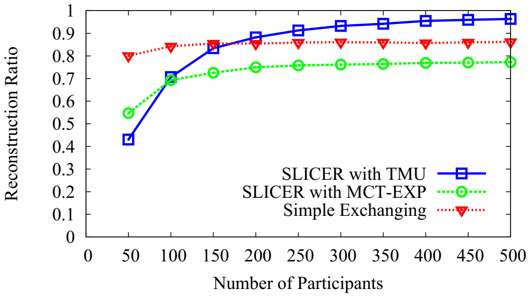

Fig. 4. Impact of Participant Number on Reconstruction Ratio

Fig. 4 shows the reconstruction ratios achieved by the three schemes with the growth of number of participants. We set the probability of slice loss to 0.2 in this simulation. To be fair, we let the three evaluated schemes have the same communication overhead, and then compare their achieved reconstruction radios. We can see from Fig. 4 that *SLICER with TMU*performs better than *Simple Exchanging* when there are sufficient number of participants (*i.e.*, >200 participants). This is because SLICER inherits high loss tolerance from erasure coding. Specifically, the reconstruction ratio of *SLICER with TMU* reaches 0.97 when there are 400 participants or more. In contrast, *Simple Exchanging* has relatively stable reconstruction ratio (about 0.86). However, we can see that *SLICER with MCT-EXP* performs not well, due to the fact that the MCT-EXP strategy may not guarantee the probability of meeting $m-1$ participants at a high level. In addition, when the number of participants is less than 150, *Simple Exchanging* performs the best. This is because *Simple Exchanging* only needs one other participant to deliver the sensing record, while SLICER needs $m-1$ participants. So *SLICER with TMU*is preferred when there are sufficient number of participants in the system.

Next, we evaluate the communication overhead of the three schemes to achieve a reconstruction ratio of 0.99, under different slice losing probabilities. We set the sensing record size to $1MB$. Three loss probabilities are evaluated. To achieve the reconstruction ratio of 0.99, the coding rate of SLICER needs to reach 10/18, 10/21, and 10/26, when the loss probability is 0.2, 0.3, and 0.4, respectively. Similarly, we also set proper transmission redundancies for the *Simple Exchanging* for different loss probabilities. From Fig. 5, we can see that the communication overhead of SLICER is always lower than *Simple Exchanging* under different losing probabilities, showing that SLICER has better loss tolerance. Although the communication overheads of three schemes increase with the loss probability, the growth speed of SLICER is much slower.

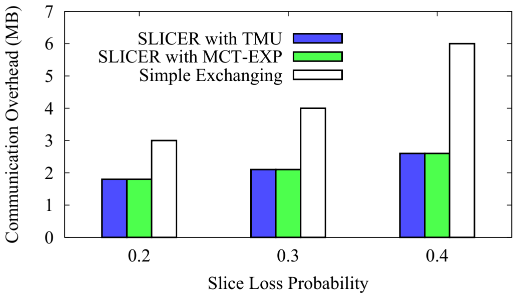

Fig. 5. Communication Overhead to Achieve Reconstruction Ratio of 0.99

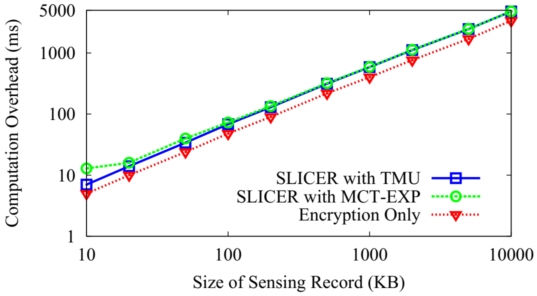

Fig. 6. Computation Overhead

We also evaluate the computation overhead of SLICER with different transfer strategies (as shown in Fig. 6, which is a log-log scale plot), comparing with the traditional encryption only scheme (RSA is adopted in this simulation and labeled with *Encryption Only*). What we consider in SLICER are only the computations needed at mobile device side, including erasure coding (Reed-Solomon [24]), hashing (SHA-1 [26]), encryption (RSA [31]), and running the two different transfer strategies. Our scheme is evaluated in windows OS environments, with C++ programmed simulator running on a computer with CPU speed of 2.40GHz. In Fig. 6, we can see that SLICER induces some extra computation overhead compared with the *Encryption Only* method, when dealing with the same size of data. This is caused mainly by the usage of erasure coding. However, the extra computation overhead is relatively small. For instance, *SLICER with TMU*, *SLICER with MCT-EXP*consumes 42.5% and 48.3% more time than the *Encryption Only* method when dealing with $10000KB$ sensing record, respectively.

Finally, we compare the total slice transfer cost when using the two strategies proposed in section III. The transfer cost on each participant is generated randomly from 0 to 1, and the meeting probability (used in Minimal Cost Transfer) comes from the statistics for 500 participants' history. The coding rate is set to 10/20. As shown in Fig. 7, *SLICER with TMU* has a higher total transfer cost, which is close to the

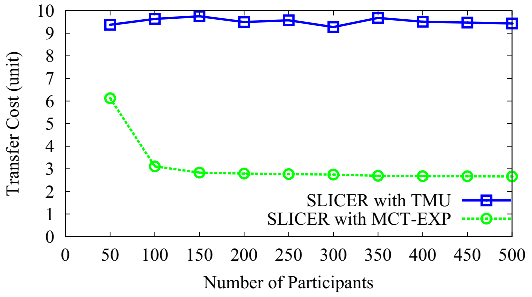

Fig. 7. Transfer Cost of Different Strategies

excepted value (*i.e.*, $0.5 \times 19 = 9.5$). Minimal Cost Transfer performs obviously better than the former one due to the well designed algorithms, especially the *MCT-EXP*. In addition, the transfer cost of SLICER converges with the growth of number of participants, because more participants will provide more meeting opportunities, higher meeting probability, and more low-cost relays to select.The results of this simulation confirm that our algorithm for minimal cost transfer can save transfer cost.

V. Related Works

In the current state-of-the-art, a number of privacy preserving techniques for participatory sensing, especially the location-based services (LBSs), have been proposed by researchers, mainly to address the privacy of data source identity, user location, trajectory, and data itself. These techniques can be classified into the following four categories.

1) Randomization Based Techniques: Randomization (noise) based technique [13], [32], [33], where noise (*e.g.*, Gaussian noise) may be added into the original data, can hide the real value of sensitive information (*e.g.*, the trend of the data over time). This method was widely studied and used in data mining field. However, the loss of data accuracy is a significant shortcoming.

2) Generalization: The k-anonymity [23] model, which aims to hide each user's sensitive information among $k-1$ others', is a universal metric for privacy preservation, and has been applied to participatory sensing in several previous works [11], [34]. However, this method usually needs an honest third-party as anonymizer, which is not allowed in ubiquitous semi-honest models. Therefore, when a more severe situation of semi-honest third-party is considered, these approaches cannot meet requirements.

3) Cloaking Techniques: Cloaking techniques usually use generalization or perturbation to replace the actual location with larger area or to cloak real location using some functions (*e.g.*, [11], [34]–[36]). However, while spatial cloaking techniques can well protect single location information, they fail to protect the trace privacy, with which user's identity is also inferable [14], [37]. Recently, several works were proposed aiming to solve the trajectory privacy problems [10], [22], [38]. However, same questions exist that the protection of privacy reduces the accuracy of reported data.

4) Cryptography Based Solutions: End-to-end encryption, which can guarantee high security of reported data, is widely used for the privacy preserving [12], [39]–[41]. However, encryption can only protect privacy from external attacks, and service provider can still easily infer users' sensitive features when encryption approach used alone. Since internal attacks are also undesirable, a scheme to prevent both external and internal attacks is quite important.

SLICER shares similarities with [22] and [12]. The former focused on the collaborative path hidden from service provider using data exchanging strategy. However, attacks from participants are not considered. The latter applied the slicing and mixing method into privacy preserving of urban sensing. However, what J. Shi *et al.* focused on were the search strategies of cover nodes, and the type of sensing data they interested was only physical data. In contrast, SLICER proposed in this paper is a k-anonymous privacy preserving scheme for participatory sensing with multimedia data, and we can achieve high data accuracy, with low communication and computation overhead.

VI. Conclusion and Future Works

In this paper, we have presented SLICER, which is a k-anonymous privacy preserving scheme for participatory sensing. SLICER integrates the technique of erasure coding and well designed slice transfer strategies, to achieve strong protection of participants' privacy as well as high data accuracy and loss tolerance, with low computation and communication overhead. For future work, one possible direction is to study the problem of privacy preservation in the query process, and design privacy preserving query schemes based on SLICER. Another possible direction is to design efficient slice transfer algorithm, considering the limitation of mobile devices' battery power, storage space, and communication bandwidth.

References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Workshop on World-Sensor-Web (WSW), co-located with ACM SenSys*, 2006.

[2] "The world in 2011: ICT Facts and Figures," International Telecommunication Union. [Online]. Available: http://www.itu.int

[3] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: a participatory sensing fuel-efficient maps application," in *MobiSys*, 2010.

[4] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *MobiSys*, 2009.

[5] M. von Kaenel, P. Sommer, and R. Wattenhofer, "Ikarus: large-scale participatory sensing at high altitudes," in *Workshop on Mobile Computing Systems and Applications*, 2011.

[6] R. K. Ganti, N. Pham, Y.-E. Tsai, and T. F. Abdelzaher, "PoolView: stream privacy for grassroots participatory sensing," in *SenSys*, 2008.

[7] J. Soldatos, M. Draief, C. Macdonald, and I. Ounis, "Multimedia search over integrated social and sensor networks," in *WWW*, 2012.

[8] "Cabspotting Project." [Online]. Available: http://cabspotting.org/

[9] Z. Yang, S. Zhong, and R. N. Wright, "Anonymity-preserving data collection," in *KDD*, 2005.

[10] J. Meyerowitz and R. Roy Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *MobiCom*, 2009.

[11] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *INFOCOM*, 2012.

[12] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: privacy-preserving data aggregation in people-centric urban sensing systems," in *INFOCOM*, 2010.

[13] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via uncertainty-aware path cloaking," in *CCS*, 2007.

[14] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao, "Privacy vulnerability of published anonymous mobility traces," in *MobiCom*, 2010.

[15] Y. Lindell and B. Pinkas, "Privacy preserving data mining," *the Journal of Cryptology*, vol. 15, no. 3, pp. 177–206, 2002.

[16] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in *KDD*, 2002.

[17] R. Wright and Z. Yang, "Privacy-preserving bayesian network structure computation on distributed heterogeneous data," in *KDD*, 2004.

[18] A. W.-C. Fu, R. C.-W. Wong, and K. Wang, "Privacy-preserving frequent pattern mining across private databases," in *ICDM*, 2005.

[19] W. Lu, A. L. Varna, and M. Wu, "Security analysis for privacy preserving search of multimedia," in *International Conference on Image Processing*, 2010.

[20] A. Adams and M. Sasse, "Taming thewolf in sheeps clothing: Privacy in multimedia communications," in *ACM Multimedia*, 1999.

[21] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, R. L. Lagendijk, and T. Toft, "Privacy-preserving face recognition," in *International Symposium on Privacy Enhancing Technologies*, 2009.

[22] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, and S. S. Kanhere, "Privacy-preserving collaborative path hiding for participatory sensing applications," in *MASS*, 2011.

[23] L. Sweeney, "k-anonymity: a model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[24] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[25] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *SIGCOMM*, 1998.

[26] D. Eastlake, 3rd and P. Jones, "US Secure Hash Algorithm 1 (SHA1)," RFC 3174, September 2001. [Online]. Available: http://www.rfc-editor.org/in-notes/rfc3174.txt

[27] G. Lin, G. Noubir, and R. Rajaraman, "Mobility models for ad hoc network simulation," in *INFOCOM*, 2004.

[28] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: a new mobility model for human walks," in *INFOCOM*, 2009.

[29] V. V. Vazirani, *Approximation Algorithms*. Springer Verlag Press, 2001.

[30] "CRAWDAD: a community resource for archiving wireless data at dartmouth." [Online]. Available: http://crawdad.cs.dartmouth.edu/

[31] R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[32] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *SIGMOD*, 2000.

[33] F. Zhang, L. He, W. He, and X. Liu, "Data perturbation with state-dependent noise for participatory sensing," in *INFOCOM*, 2012.

[34] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Computer Communications*, vol. 33, no. 11, pp. 1266–1280, July 2010.

[35] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson, "Virtual trip lines for distributed privacy-preserving traffic monitoring," in *MobiSys*, 2008.

[36] Y. Wang, D. Xu, X. He, C. Zhang, F. Li, and B. Xu, "L2P2: location-aware location privacy protection for location-based services," in *INFOCOM*, 2012.

[37] H. Zang and J. Bolot, "Anonymization of location data does not work: a large-scale measurement study," in *MobiCom*, 2011.

[38] T. Xu and Y. Cai, "Feeling-based location privacy protection for location-based services," in *CCS*, 2009.

[39] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *SIGMOD*, 2010.

[40] J. Manweiler, R. Scudellari, and L. P. Cox, "SMILE: encounter-based trust for mobile social services," in *CCS*, 2009.

[41] M. Bechler, H.-J. Hof, D. Kraft, F. Pählke, and L. Wolf, "A cluster-based security architecture for ad hoc networks," in *INFOCOM*, 2004.