Delphine Christin, Julien Guillemet, Andreas Reinhardt, Matthias Hollick, Salil S. Kanhere: Privacy-preserving Collaborative Path Hiding for Participatory Sensing Applications. In: Proceedings of the 8th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS), pp. 341–350, 2011.

Privacy-preserving Collaborative Path Hiding for Participatory Sensing Applications

Delphine Christin*, Julien Guillemet*, Andreas Reinhardt†, Matthias Hollick*, Salil S. Kanhere‡
* Secure Mobile Networking Lab, Technische Universität Darmstadt, Darmstadt, Germany
† Multimedia Communications Lab, Technische Universität Darmstadt, Darmstadt, Germany
‡ School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
Email: *delphine.christin@seemoo.tu-darmstadt.de*

***Abstract*—The presence of multimodal sensors on current mobile phones enables a broad range of novel mobile applications including, e.g., monitoring noise pollution or traffic and road conditions in urban environments. Data of unprecedented quantity and quality can be collected and reported by a possible user base of billions of mobile phone subscribers worldwide. The collection of detailed sensor and location data may however compromise user privacy. In this paper, we present a decentralized mechanism to preserve location privacy during the collection of sensor readings. As most sensor readings are geotagged, we propose to exchange them between users in physical proximity in order to jumble the paths followed by the users. We evaluate different strategies to exchange and report the sensor readings to the application using real-world GPS traces of mobile users. The results demonstrate the feasibility and efficacy of our proposed scheme, which can obfuscate up to 100% of the visited locations in the best instances.**

I. Introduction

Recent mobile phones open novel perspectives in terms of sensing. In addition to widespread wireless technologies (e.g., Wi-Fi, 3G, or Bluetooth), they are equipped with a plethora of embedded sensors (e.g., microphone, camera, accelerometer, and gyroscope) as well as advanced processing and storage capabilities. With an estimated number of 5 billion users worldwide [1], they offer an unprecedented spatial coverage and a well-established communication infrastructure virtually diminishing the deployment costs to zero. The utilization of mobile phones as sensing devices is referred to as *participatory sensing* ([2], [3]) and has opened the doors to the development of a plethora of applications including the collection and sharing of information about personal diets [4], urban noise pollution [5], and cyclist experiences [6].

In virtually all of these applications, the sensor readings such as images or sound samples are tagged with the corresponding location coordinates and time and are uploaded to a central server, which is typically controlled by the application. In absence of any protection mechanism, this spatiotemporal information may leak privacy-sensitive information about the participants by revealing, e.g., the paths followed by the participants, their routines and habits, as well as their home and workplace locations [7]. In the face of such significant threats to their privacy, persuading participants to willingly contribute data would be next to impossible. Lack of sufficient participants would in turn minimize the benefit of most participatory sensing applications. Mechanisms preserving the location privacy of the participants are therefore essential to encourage contribution and gain widespread acceptance among participants.

Our contribution is as follows. We present a decentralized and collaborative mechanism to preserve the location privacy of participants and, more particularly, the paths they follow during the collection of sensor readings. We propose to exchange the collected sensor readings between participants who physically meet. By exchanging their sensor readings, the participants jumble their paths; the prior path of one participant becomes that of another participant and vice versa. The repetition of the jumbling process at each encounter results in the construction of paths composed of concatenated subpaths from multiple participants. As a result, the sensor readings, which are reported to the server by the participants, do not disclose the actual paths, but instead a path jumbled with other participants. The strategies for exchanging and reporting the sensor readings can be adjusted to the desired level of privacy and reporting latency. With our approach, the participants thus collaborate to cover their tracks and preserve their privacy in a decentralized fashion.

This paper is structured as follows. In Section II, we define our system and adversary models. We present our concept to protect the location privacy of the participants in Section III. We analyze the design space by considering different strategies to exchange the sensor readings between the participants in Section IV as well as introducing strategies to report the sensor readings to the server in Section V. We define design criteria for both exchange and reporting strategies, which serve as basis for our design decisions. Based on this design space analysis, we present different variants of exchange and reporting strategies. Moreover, we evaluate these strategies using real-world GPS traces of human mobility. We measure the impact of different strategy combinations on the privacy of the participants in Section VI. The results demonstrate the feasibility of our concept under realistic user mobility assumptions and provide insights about the dependencies between jumbling efficacy, reporting latency and privacy protection. Finally, we discuss our results in Section VII and survey related work in Section VIII, before concluding this paper in Section IX.

II. Assumptions and Models

A. System Model

For our system, we assume participatory sensing applications without real-time constraints for data delivery. Examples include monitoring noise pollution [5] and road conditions [8] as well as documenting personal diets [4]. In these applications, the participants carry mobile phones equipped with embedded sensors, wireless interfaces and positioning systems (e.g., GPS, Wi-Fi, or cellular network based triangulation [9]). The mobile phones autonomously collect sensor readings (e.g., sound samples, pictures, and accelerometer data). However, the participants can control the activation/deactivation of the sensing function and its sampling period. Each sensor reading is stamped with the collection time and location information to form the following triplet $T =< t, l, s >$ with t: time, l: location, s: sensor reading, e.g., present in the form of vectors (like 3-axis accelerometer readings) or scalar value (such as noise level in decibels). Additional processing can be locally applied on the sensor readings to extract features and/or avoid the disclosure of sensitive information. For example, possibilities include extracting the noise level from the collected sound samples or obfuscating recorded conversations in applications monitoring noise pollution.

The triplets are then autonomously reported to the application server. We assume that the participants can configure reporting settings including reporting frequency, reporting locations, or when they have access to a free Wi-Fi connection. The application server is able to establish a link between the reported triplets and the participants who reported them. The establishment of this link can be based on either explicit identifiers, such as user ID and pseudonyms, or the analysis of reporting metadata to e.g. infer the location from the used IP addresses. The application can utilize the established links to assign reputation to the participants [10], as from the application perspective the triplets are assumed to be reported by the participants who collect them.

Finally, the application analyzes the reported triplets to build summary maps (e.g., illustrating the noise level or the road conditions across the city), or provide statistics (e.g., time and frequency of meals during the diet program). The results can be accessible to the public, groups of participants, or only the participant himself in the case of personal sensing application such as DietSense [4].

B. Adversary Model

We define as adversary each party who aims at gaining access to the locations visited and the paths followed by the participants. The motivations of the adversary include simple curiosity and willingness to harm or commercially exploit the disclosed information. We identify two major categories of adversary: Malicious application administrators and malicious participants. Note that the latter category is an artifact of the collaborative nature of our approach.

Application administrators represent a common threat to privacy in participatory sensing deployments. They have direct access to the data reported by the participants and stored on the application server. In absence of privacy-preserving mechanisms locally applied on the mobile phones, the triplets contain information about the visited locations and disclose the paths followed by the participants that may provide insights about the participants' lifestyle. For example, the sensor readings collected while participants are walking from their offices to their homes might reveal their exact paths as well as their start and end locations. The participants are therefore bound to trust the administrators to neither misuse their privacy-sensitive information nor to disclose it to untrusted parties.

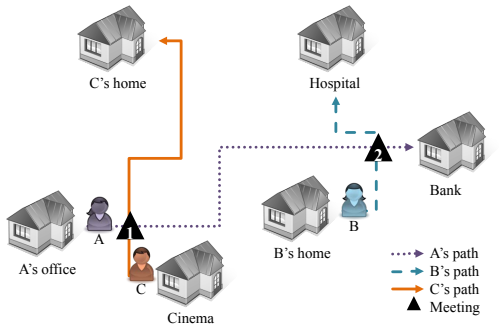
In our approach, malicious participants can also become adversaries. For example, they can develop strategies to position themselves in crowded areas to meet as many participants as possible or act as a repeater exchanging triplets from one participant with another one. To counter such threats, we design specific exchange strategies, which are presented in Section IV. Using these privacy-aware exchange strategies, the risk of disclosing information to malicious participants can be minimized, although not completely eliminated due to the cooperative nature of our concept.

III. Path Jumbling Concept

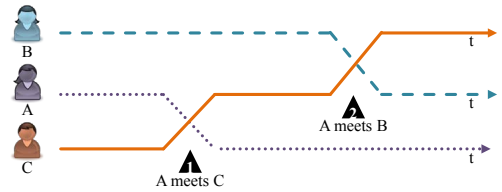
The objective of our concept is to break the link between the spatiotemporal context (i.e., time and location) at which the sensor readings were taken and the identity of the participants (i.e., mobile devices) in order to protect their privacy. The spatiotemporal context reveals the visited locations and paths followed by the participants during the sensing process, thus providing insights about the participants' lifestyles. In our decentralized approach, the participants collaborate to protect their privacy. As the triplets contain the spatiotemporal information, triplets collected on user devices between the participants in physical proximity are jumbled and thus unlinked from their original collectors.

To illustrate our concept, we consider the example in Fig. 1. We assume that the participants A, B, and C follow the paths illustrated in Fig. 1(a) while they collect triplets and meet according to the timeline represented in Fig. 1(b). In this example, A first meets C and exchanges his previously collected triplets with him. The selection of the triplets is determined by the exchange strategy. Here, we assume that A and C exchange all triplets they collected up to the moment of their encounter. Note that different exchange strategies can be envisaged, as discussed in Section IV. After their meeting, A and C continue to collect triplets while they are continuing along their routes. When A meets B, both exchange their triplets (including those already exchanged by A) according to the exchange strategy. After their meeting, they collect triplets until reaching their final destination where they terminate the sensing process.

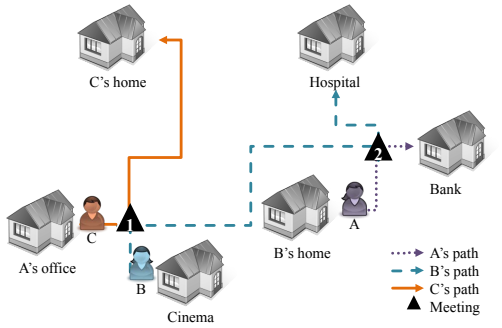
In this example, we assume that the participants configure their reporting strategy to daily upload the triplets when the sensing process is terminated. Note that the participants can select other reporting strategies detailed in Section V, which have an impact on the achievable privacy protection.



(a) Actual paths



(b) Meeting and exchange timeline



(c) Jumbled paths

Fig. 1. Meetings and exchanges of the participants A, B, and C

By applying our mechanism, the participants report triplets partially collected by themselves and the ones exchanged with other participants to the application server. Based on these triplets, the application assumes that participant A walked from B's home to the bank. The participant B traveled from the cinema to the hospital, and the participant traveled C from A's office to his home, as depicted in Fig. 1(c). As the actual paths followed by the participants are mostly hidden to the server, the privacy of the participants is respected.

IV. Exchange Strategies

As mentioned in Section III, different exchange strategies can be applied to jumble the triplets between the participants. In this section, we analyze the design space and select different strategies to investigate their impact on the performance of our concept in Section VI.

A. Design Space Analysis

Our analysis primarily concentrates on the selection modality of the triplets to exchange. This includes a discussion about the number and the selection of triplets, both locally collected and received from other participants, that will be exchanged. We consider the following design alternatives:

Partial vs. complete exchanges: The triplets can be exchanged either partially or completely. Opposed to the traces resulting from the partial exchange of triplets, paths generated by the latter alternative are realistic (see example in Section III), as if captured by a single real person. When such paths are reported to the application after the jumbling process, the participants' actual traces are obfuscated by realistic and coherent substitutes, and thereby cater for privacy protection against malicious application administrators. However, this strategy discloses the path followed by each participant between two successive encounters to other participants. In case of frequent meetings, the disclosed path may be limited to a small distance that malicious participants may also visually observe. In comparison, partial exchanges diminish the amount of information disclosed to other participants. Depending on their selection, the exchanged triplets may however still contain sensitive information. Besides, the participants will report a higher percentage of own triplets to the server, as they only partially exchange them. As a result, they will reveal more information to the server about themselves than with complete exchanges.

Individual vs. consecutive triplets: If the aforementioned partial strategy is applied, the triplets to exchange can be chosen either randomly, such that they represent a collection of sparse locations, or by selecting coherent path segments based on consecutive triplets. Depending on the selection and the length of the exchanged path segments, however, participants potentially disclose sensitive information to other users. When malicious users are present, the exchange of individual triplets represents a more secure approach, as it only provides disjoint spatiotemporal information to other participants. As a downside, however, the reported triplets may then form unrealistic paths, e.g., improbable covered distances between consecutive triplets, which can be easily detected by the application server and its administrators.

Symmetric vs. asymmetric exchanges: Another design dimension to explore is the reciprocity in the amount of triplets exchanged during an encounter. A symmetric exchange supports the collaborative nature of our concept, as the participants benefit from similar exchange conditions and potential reporting overhead can be distributed between them in a fair manner. Instead of choosing a predetermined value for all participants, a negotiation phase is used before the actual data transmission to agree on the amount of exchanged triplets. However, this negotiation must take into consideration that malicious participants may be willing to exchange a large amount of data with other participants to gather as much sensitive information as possible. In comparison, an asymmetric exchange allows each participant to individually determine how many triplets he exchanges, implying that participants can receive more triplets than they exchange. These participants will thus report more triplets to the server introducing additional overhead for them.

The above discussions highlight that each alternative presents advantages as well as drawbacks with regard to potential information disclosed to the server and the other participants, or introduced overhead. Tradeoffs between advantages and drawbacks must thus be found.

B. Design Decisions

Based on the results of the design space analysis, we select the following exchange strategies for our further analysis, which cover different combinations of the above design alternatives:

Realistic exchange strategy: The participants exchange their entire set of collected/exchanged triplets at each meeting. Assuming that the participants A and C collected 6 and 3 triplets before their meeting, A receives the 3 triplets from C, and C receives the 6 triplets from A, as illustrated in Fig. 2(a). This strategy exchanges consecutive triplets, which form realistic path segments (see Section III). The exchange can be asymmetric, if the amounts of collected/exchanged triplets differ between both participants.

Random-unfair exchange strategy: Each participant independently and randomly determines the amount of triplets he wants to exchange. In Fig. 2(b), A exchanges 5 of 6 triplets, while C exchanges 2 of 3 triplets. Moreover, each triplet to exchange is selected randomly. In our example, A selects the triplets A_1, A_2, A_3, A_4, and A_6, whereas C selects the triplets C_1 and C_3. In comparison to the realistic strategy, the exchanges are mostly partial and involve triplets selected individually. The exchanges can be asymmetric, depending on the amount of exchanged triplets.

Random-fair exchange strategy: The participants agree on a common amount of n triplets to exchange at each meeting. Each participant advertises the amount of triplets available for exchange. In Fig. 2(c), A and C advertise 6 triplets and 3 triplets, respectively. The amount of exchanged triplets is determined randomly between 1 and the minimum of both advertised values. In our example, a value of 1 triplet is chosen. Then, A and C randomly select the triplet to exchange. In comparison with the random-unfair strategy, the fair variant ensures the symmetry of each exchange that fairly jumbles the triplets and equally distributes the reporting overhead between the participants.

Furthermore, we complete these strategies by introducing additional features to prevent consecutive exchanges with the same participants and unidirectional exchanges of triplets. The first feature only allows participants to exchange triplets with the same participant when they have exchanged data with at least x other participants between two encounters. A malicious participant can thus not easily recover a full collection of triplets by simply following targeted participants and constantly exchanging triplets with them. Without this feature, already exchanged triplets may be exchanged again and may come back to the participants who collected them, lowering the jumbling degree of the triplets and thus the benefits of our approach. Moreover, we introduce a tit-for-tat mechanism where triplets are exchanged alternatively to ensure that malicious participants cannot receive an entire set of triplets without providing any triplet in exchange. Additionally, both participants have an equivalent opportunity to exchange triplets even in case of early abortion of the exchange due to technical reasons or divergent user mobility.

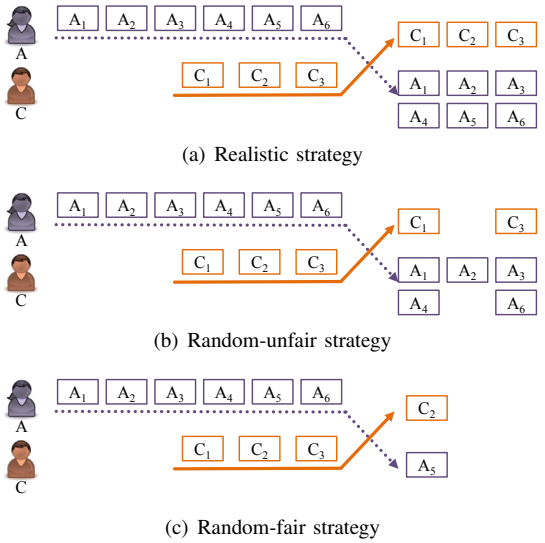


Fig. 2. Comparison of the exchange strategies

V. Reporting Strategies

In addition to the exchange strategy, different variants can be envisaged to report the triplets to the application server. In this section, we analyze and discuss design alternatives before presenting our design decisions.

A. Design Space Analysis

The privacy protection provided by our approach depends on the meeting pattern of the participants. In extreme cases, participants may not be able to exchange triplets with other participants during long time periods. The non-jumbled triplets can be either reported to the application server or stored until the next meeting. While reporting these triplets to the server would reveal the original paths followed by the participants and hence, breach their privacy, waiting until the next meeting to report the triplets would introduce additional delays for the application. Within the scope of this analysis, we investigate the tradeoff between privacy and latency by considering the following reporting strategies:

Time-based strategy: The triplets are periodically reported to the server. The application is thus ensured to timely receive triplets. However, the triplets can be reported without having been jumbled in absence of encounters during the considered period.

Exchange-based strategy: The triplets are reported to the server after each meeting. The reporting latency is thus determined by the frequency of the meetings. Although this strategy ensures that the triplets have been jumbled once before their report, no guarantee is provided on the achieved degree

of privacy/jumbling. For example, only one of the reported triplets may have been jumbled with another participant.

Metric-based strategy: The triplets are only reported to the server after reaching privacy-related thresholds, such as a given number of triplet exchanges. This strategy thus guarantees the participants that their privacy is respected to a degree defined by the threshold. However, these thresholds may increase the latency between two reports, as multiple meetings may be necessary to reach them.

The discussed reporting alternatives highlight that high privacy protection and low latency for the application are difficult to combine. A tradeoff must thus be found between both parameters, especially as participants may refuse to contribute to the application if their privacy is not protected. To the interests of the application, these contributions must be encouraged, meaning that strategies with guaranteed privacy should be preferred, despite the introduction of additional latency for the application.

B. Design Decisions

We select different combinations of the previously discussed reporting alternatives in order to investigate their actual impact on the privacy protection in our evaluation. Note that the parameterization of the strategies has been influenced by the real-world dataset used for the evaluation (see Section VI).

Hourly and daily strategies: The triplets are reported hourly and daily, respectively. Both time-based strategies may report non-jumbled triplets in absence of meetings. In comparison with the hourly strategy, the daily strategy offers a longer period during which meetings may occur.

1-Exchange strategy: The triplets are reported after each exchange. We introduce a random waiting period before the upload to prevent the server from identifying the participants who exchanged their triplets by analyzing simultaneous reports.

Jumbling-based strategy: The triplets are reported if the percentage of jumbled triplets (i.e., collected by others) reaches a given threshold. We choose the following thresholds: 25%, 50%, and 75%. This metric-based strategy allows controlling that the triplets/paths have been sufficiently jumbled to provide a certain privacy guarantee. A high percentage indicates that only few triplets were collected by the participants themselves and their reporting discloses thus little information about the paths actually followed by them.

Distance-based strategy: The triplets are reported if the average distance between each location of the actual and jumbled paths is above a given threshold. We select three thresholds: 1km, 2km, and 5km. Similarly to the former strategy, it provides an estimation of the path privacy protection. If the distance is small, the jumbled path remains in proximity of the actual path and may still contain sensitive locations. If the distance is large, fewer insights about the participants can be inferred.

VI. Evaluation

In this section, we describe the performance evaluation of the selected exchange and reporting strategies detailed in Section IV-B and Section V-B, respectively. We first discuss the characteristic properties of the utilized dataset before describing the settings and the metrics introduced for our evaluation. Finally, we present the results of our evaluation and highlight particular findings.

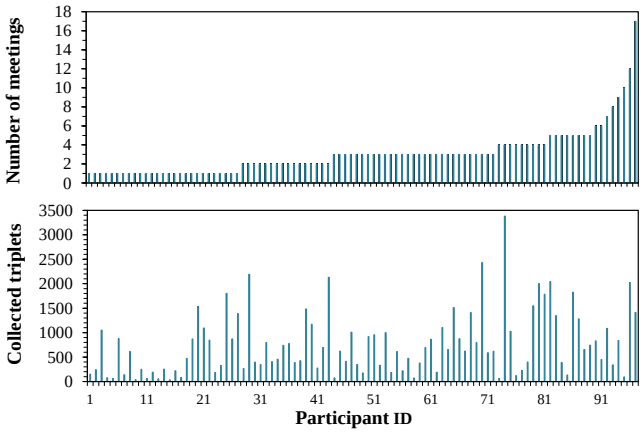


Fig. 3. Number of meetings and collected triplets per participant

A. Dataset

Our evaluation is based on the GPS traces from the GeoLife project ([11], [12]). In this real-world deployment, the participants carried GPS-enabled devices to monitor their location. We extend the initial scenario to a participatory sensing application by assuming that a triplet was collected at each monitored location.

We selected 97 participants having at least met one other participant, and we observed their mobility and meeting pattern during 24 hours. The meeting distribution of the selected participants is depicted in the upper part of Fig. 3, while their respective number of collected triplets is presented in the lower part of Fig. 3. Note that the difference between the numbers of collected triplets is due to the possibility for the participants to select the collection frequency as well as activate/deactivate the sensing function.

Among the 97 participants, we selected 3 participants who represent the extreme and average cases in terms of meetings. With a single meeting, the first participant (ID=19) is designated as *worst case*, while the second participant (ID=55) and third participant (ID=97) are referred to as *mean case* and *best case* with meeting counts of 3 and 17, respectively.

B. Settings and Metrics

To evaluate our approach, we simulated the exchange and reporting of triplets based on the GPS traces. When the participants were in physical proximity, they exchanged their triplets according to one of the selected exchange strategies (see Section IV-B). To avoid constant exchanges between the same participants, each participant was only able to exchange with the same participant again, if he met three other participants in between. Although this value has been selected based on the characteristics of the dataset, it could also be easily

determined dynamically in a real-world deployment. Besides, the participants reported the triplets to the application server with one of the selected reporting strategies (see Section V-B). We therefore conducted a cross-evaluation of both exchange and reporting strategies, where each possible combination of strategies is evaluated using the following metrics:

Jumbling degree: It measures the average percentage of reported triplets having been jumbled with other participants. A high percentage thus indicates that only few triplets collected by the participants themselves are reported meaning that little information about the participants' paths is disclosed. This metric provides thus insights about the level of obfuscation achieved at the time of the reporting to the server.

Distance: It estimates the average distance between the actual path followed by the participants and the jumbled path resulting from the exchange. The metric provides an estimation of the path privacy protection. A small distance indicates that the reported path remains in proximity of the actual path. The reported path may thus still reveal sensitive locations visited by the participants.

Overhead: It compares the average amount of triplets having been reported after jumbling with the amount of triplets having been collected. It thus measures the reporting overhead caused by our approach. A positive factor/percentage means that the participants need to report more triplets than they collected themselves after applying our mechanism.

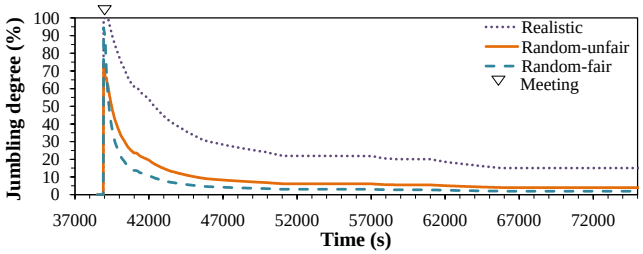
C. Results

In this section, we compare and evaluate the impact of the selected exchange and reporting strategies on the following metrics defined in Section VI-B:

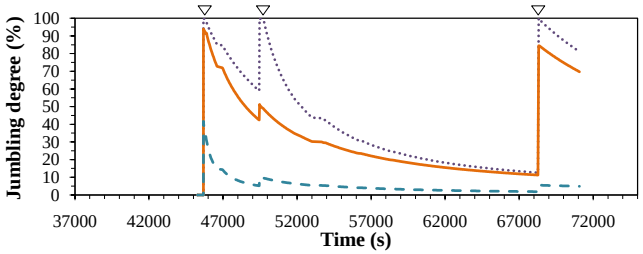
1) Jumbling degree: We consider the realistic, random-fair, and random-unfair exchange strategies and assume first that the triplets are reported at the end of the sensing process. We examine three different cases: Worst, mean, and best cases corresponding to the participants introduced in Section VI-A. Fig. 4 illustrates the temporal evolution of the jumbling degree for each participant. Note that the participants present different sensing periods, as they were able to activate/deactivate the sensing function. The figures highlight that the realistic strategy ensures the highest jumbling degree, as all triplets are exchanged at each meeting. In comparison, both random-unfair and random-fair strategies reach lower jumbling degrees due to partial exchanges of randomly selected triplets. Moreover, the random-unfair strategy allows a higher jumbling degree than the random-fair one. However, the jumbling degrees of both random-fair and random-unfair strategies primarily depend on the generated random selection determining the amount of triplets to exchange.

We next analyze the combination of exchange and reporting strategies for the entire dataset. For each combination, we calculate the minimal, mean, and maximal jumbling degree of the triplets reported to the application server. Table I shows the obtained results, which can be summarized as follows:

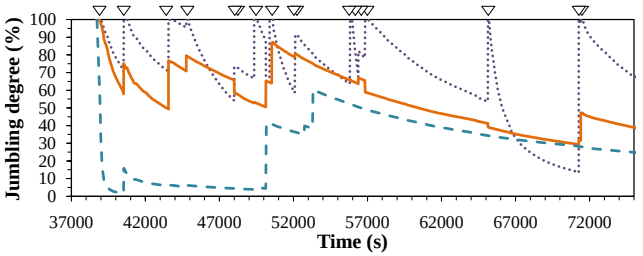
The *realistic strategy* combined with all reporting strategies shows a higher jumbling degree than the random-unfair and



(a) Worst case (participant with one meeting)



(b) Mean case (participant with three meetings)



(c) Best case (participant with 17 meetings)

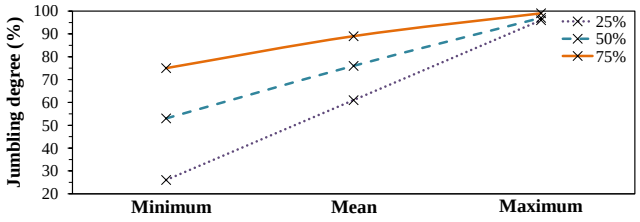
Fig. 4. Jumbling degree over time for the three selected participants

random-fair strategies. Except for the time-based reporting strategies, the jumbling degree reaches 100%, meaning that all triplets were jumbled before their reporting to the application server, even for participants with a single meeting only. This implies that the paths are protected independently of the selected reporting strategy and even if the participants meet only once. In addition to high jumbling degree, this exchange strategy allows the application to timely receive the triplets, as the thresholds triggering the report can be reached after a unique meeting, while multiple meetings may be needed in the case of the random-unfair and random-fair strategies.

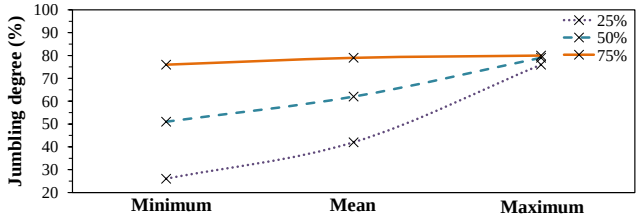
The *random-fair strategy* obtains in general lower jumbling degrees than the *random-unfair strategy* due to its fairness constraint, as illustrated in Fig. 5. For both random-unfair and random-fair strategies, the jumbling-based reporting strategies show higher jumbling degrees than with other reporting strategies. Note that the minimum values of the jumbling degree are at least equal to the threshold of the applied jumbling-based strategy. For example, the minimal jumbling degree obtained by applying the random-unfair exchange strategy and the 25%-variant jumbling-based reporting strategy is equal to at least 25%. The jumbling-based reporting strategies therefore guarantee minimal jumbling degrees, which is not the case for the other reporting strategies. The time-based reporting

TABLE I
STRATEGY COMPARISON: JUMBLING DEGREE, DISTANCE, AND OVERHEAD

			Metrics				
			Jumbling degree (%)			Distance (km)	Overhead (%)
Exchange strategy	Reporting strategy	Threshold	Minimum	Mean	Maximum	Median	Median
Realistic	Time-based	hourly	0	17	49	2	8
		daily		59	100	5	7
	Exchange-based	1	100	100		4	-8
	Jumbling-based	25%					
		50%					
		75%					
	Distance-based	1 km				5	0
		2 km					-6
		5 km				8	-16
Random-unfair	Time-based	hourly	0	8	32	2	5
		daily		28	88	5	7
	Exchange-based	1		36	97	4	-4
	Jumbling-based	25%	26	61	96	5	-26
		50%	53	76	97		-46
		75%	75	89	99		-78
	Distance-based	1 km	0	36	97		-3
		2 km		30	96		-9
		5 km	1	49	98	9	-27
Random-fair	Time-based	hourly	0	2	11	2	0
		daily		10	39	5	
	Exchange-based	1		22	65	4	
	Jumbling-based	25%	26	42	76	6	
		50%	51	62	79	5	
		75%	76	79	80	9	
	Distance-based	1 km	0	21	65	7	
		2 km		19	63	4	
		5 km	1	21	80	7	



(a) Random-unfair strategy



(b) Random-fair strategy

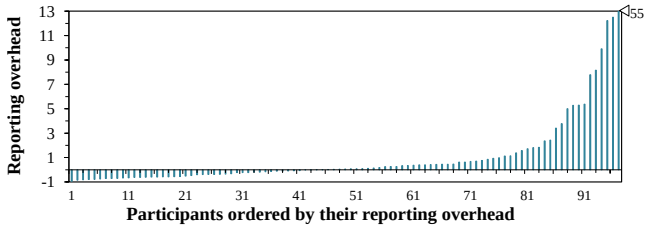
Fig. 5. Jumbling degree of the jumbling-based reporting strategies

strategies show particularly low jumbling degree and do not provide any guarantee, as the triplets are reported even in absence of meetings. All other reporting strategies achieve high jumbling degrees.

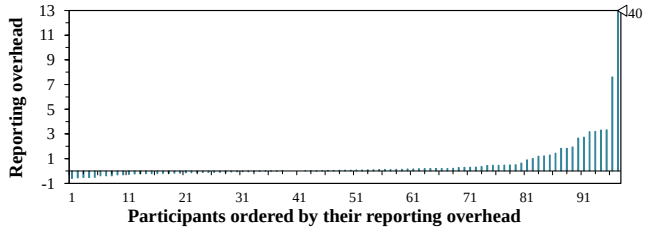
2) Distance: Similarly to the evaluation of the jumbling degree, we consider each combination of exchange and reporting strategies and calculate the distance between the collected and the reported triplets for all participants. The second column to the right of Table I presents the obtained median distances. Note that the obtained values depend on the mobility pattern of the participants. In isolation, these values only provide limited insights, while their comparison allows to evaluate the performance of the combined exchange and reporting strategies for preserving the path privacy. In average over all reporting strategies, the median distance between positions on the real and jumbled path obtained by the realistic strategy is 4 km, while it is 5 km for both random-unfair and random-fair strategies. The impact of the different exchange strategies is thus only slightly distinguishable. The analysis of the different reporting strategies shows that there is no major difference in terms of distance between them. However, the distance-based reporting strategies ensure that the reported triplets at least reach their respective distance threshold. The participants are therefore guaranteed that the reported triplets present a distance to the actual path greater than the predetermined threshold, and hence, that their privacy is respected. The hourly strategy performs poorly, as the triplets are reported even in absence of meetings.

3) Overhead: Except for the random-fair strategy, the exchange strategies can influence the amount of triplets to report, as the participants can receive a lower, equal, or higher amount of triplets than they provide. The difference can thus increase, leave unchanged, or reduce the reporting overhead.

In a first step, we evaluate the potential reporting overhead introduced by the exchange strategies combined with the daily



(a) Realistic exchange strategy



(b) Random-unfair exchange strategy

Fig. 6. Distribution of the reporting overhead factor for the daily reporting strategy

reporting strategy. Fig. 6 illustrates the distribution of the reporting overhead factor. The realistic strategy introduces a greater overhead factor than the random-unfair strategy. The overhead remains negative for 44 participants with realistic strategy and 40 participants with the random-unfair strategy. This implies that these participants report fewer triplets than collected by themselves with our approach. On the other hand, the remaining participants report more triplets than they collected with a factor reaching up to 55 for the realistic strategy and 40 for random-fair strategy.

Secondly, we examine the median reporting overhead introduced by the exchange strategies in combination with different reporting strategies. The results are summarized in the third column to the right of Table I. By design, the random-fair exchange strategy does not introduce any overhead. The comparison of both realistic and random-unfair strategies shows that the overhead median is -4% on average for the realistic strategy and -20% on average for the random-unfair strategies. This result confirms and extends the prior finding: The overhead introduced by the realistic strategy is generally higher than with the random-unfair strategy. The overhead difference however depends on the random value, determining the amount of triplets to exchange in the random-fair strategy. Moreover, the reporting performance is also influenced by the random selection of the triplets, as the thresholds can be either exceeded or not, depending on the selected triplets. Note that the median reporting overhead is negative in most cases. Intuitively, we would expect its value to globally be zero, as the triplets given by one participant are received by another. However, we only consider the triplets that are actually reported (and not only exchanged) in the table. If the reporting conditions are not fulfilled, the triplets are not reported and in consequence not taken into account in the calculation of the reporting overhead.

VII. Discussion

The above evaluation shows that the realistic exchange strategy provides the best results in terms of their jumbling degree. Except for the time-based reporting strategy, the jumbling degree reaches 100% even after a single meeting, while additional meetings increase the mixing of the triplets among the participants. As a single meeting is sufficient to reach the reporting threshold, the latency between two reports exclusively depends on the meeting pattern. Moreover, the participants are ensured that their triplets are sufficiently jumbled before reporting. However, this strategy requires a high degree of trust in other participants and introduces substantial overhead, which is not fairly distributed among the users. We observe extreme cases where participants with thousands of triplets meet participants with few triplets, causing a high overhead for the latter users. In comparison, the random-unfair exchange strategy requires a lower degree of trust and introduces less overhead. However, the partial exchanges of triplets diminishes the jumbling degree and can delay the reporting, as multiple meetings can be necessary to reach the report thresholds. The latency between two reports not only depends on the meeting pattern, but also on the random selection of the triplets. The overhead is also not equally distributed among the participants. Finally, the random-fair strategy presents the most restrictive exchange condition, as the amount of triplets to exchange is randomly determined based on the minimum amount of triplets collected by the participants. The required degree of trust in other participants is the lowest and there is no reporting overhead for any participant. However, the restrictive exchange condition limits the amount of exchanged triplets that reduces the jumbling degree and can increase the latency between two reports, as the thresholds can be more difficult to reach.

The obtained results depend on the characteristics of the utilized dataset. While our evaluation cannot be generalized to all scenarios, it shows the feasibility of our approach under realistic assumptions. We consequently plan to integrate our approach in a real-world deployment to further investigate its performances and practical issues. For this integration, we plan to the application of the AllJoyn technology [13], which enables ad hoc communication between devices without any user interaction. The participants simply broadcast messages quoting the amount of triplets they want to exchange. When participants discover other users contributing to the same application, they establish a handshake and start to alternatively exchange the triplets using the tit-for-tat mechanism. The communication between both devices is protected against eavesdropping and data manipulation by the SSL/TLS protocol supported by AllJoyn.

Furthermore, we analyze in our evaluation the performance of our concept in isolation. However, our approach can be combined with additional mechanisms to further increase the degree of privacy protection. For example, we envision that the participants can configure settings defining spatiotemporal zones, in which no sensor reading is captured (e.g., after 6 pm

in a range of 2 km around their homes), select the granularity of the released location information (e.g., zip code instead of precise coordinates), or review the collected triplets before their report to ensure that no sensitive location is disclosed. However, the latter aspect negatively impacts the usability, as it requires frequent interventions of the participants.

Moreover, we considered in our evaluation exchanges involving two participants due to the sparse participant distribution in the dataset. However, our approach can be easily extended to cover meetings involving multiple participants by introducing,e.g., round-robin algorithms.

VIII. Related Work

In this paper, we present an approach to protect the privacy of the participants by breaking the link between the spatiotemporal context of the sensor readings and the participants who collect them. A simple alternative to our mechanism could be that the participants use pseudonyms to report the triplets to the server. However, it was demonstrated in [14], that the application can infer the real identity of the participants by tracking their location traces over multiple reports, as it may expose the location of their workplaces and homes.

In the current state-of-the-art, additional measures tailored to participatory sensing applications and preserving location privacy include spatial cloaking and data perturbation. Spatial cloaking builds groups of participants that share a common attribute (e.g., k participants located in the same district) to render them indistinguishable from each other. For example, the real location of the participants can be replaced using the averaged location of the k nearest participants [15]. However, spatial cloaking relies on a third-party entity managing the perturbation of the locations for all participants. To generate the cloaked values, the participants need to report their exact locations to the third-party entity, endangering their privacy. On the other hand, data perturbation intentionally perturbs the location traces by adding artificial noise (e.g. Gaussian noise) in order to conceal the individual traces. However, the noise model is selected by the application and influences the efficiency of the privacy protection. Indeed, independent random noise has been demonstrated insufficient to prevent adversaries from reconstructing the original data [14]. The participants must therefore trust the application to protect efficiently their privacy, while they ensure their privacy themselves in our approach.

Our approach shares features with the concept of mix zones [16], where the participants change their pseudonyms when they encounter other participants. However, our mechanism does not involve pseudonyms, but is solely based on the triplets collected by the participants, which are actually exchanged. Our concept shares also similarities with the data aggregation scheme proposed in [17]. This decentralized scheme is based on data slices equally distributed between neighbors before being reported to an aggregation server. Instead of only considering an equal distribution between neighbors, we examine multiple exchange strategies as well as reporting strategies that are not addressed in this prior work. Moreover, we investigate different dimensions in our evaluation and we do not assume the existence of an aggregation server. Our mechanism is tailored for common participatory sensing applications, where each participant individually and directly reports triplets to the application server.

IX. Conclusions

We have presented a collaborative and decentralized approach to preserve the location privacy of users contributing to participatory sensing applications. Our approach is solely based on the exchange of the collected sensor readings between participants in physical proximity in order to conceal the paths they have followed. We have examined multiple exchange strategies in combination with different reporting strategies to determine their impacts on the privacy protection. Based on the utilized realistic dataset, we have shown that the realistic exchange strategy guarantees privacy independently of the applied reporting strategy, unless time-based reporting strategies are being used. However, it requires a high degree of trust in other participants and generates a potentially large reporting overhead to some of the users. In comparison, the random-unfair and random-fair strategies require a larger amount of meetings to ensure similar guarantees, but require a lower degree of trust and generate less overhead. For the reporting strategies, we have shown that the threshold-based variants are the only ones to provide strong privacy guarantees on the reported triplets in combination with all exchange strategies. As a result of our cross-evaluation, we have provided insights for the strategy selection depending on the desired privacy guarantees, the degree of trust in other participants and the willingness to equally share the reporting overhead. Using our approach, the participants can therefore customize their exchange and reporting settings depending on their preferences and protect their location privacy themselves.

Acknowledgment

This work was supported by CASED (www.cased.de).

References

[1] International Communication Union. The World in 2010: ICT Facts and Figures. [Online]. Available: http://www.itu.int

[2] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, and R. Peterson, "People-centric Urban Sensing," in *Proc. of the 2nd Annual International Wireless Internet Conference (WICON)*, 2006.

[3] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. Srivastava, "Participatory Sensing," in *Proc. of the 1st Workshop on World-Sensor-Web (WSW)*, 2006.

[4] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen, "Image Browsing, Processing, and Clustering for Participatory Sensing: Lessons from a DietSense Prototype," in *Proc. of the 4th Workshop on Embedded Networked Sensors (EmNets)*, 2007.

[5] R. Rana, C. Chou, S. Kanhere, N. Bulusu, and W. Hu, "Ear-Phone: An End-to-end Participatory Urban Noise Mapping System," in *Proc. of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2010.

[6] S. Eisenman, E. Miluzzo, N. Lane, R. Peterson, G. Ahn, and A. Campbell, "BikeNet: A Mobile Sensing System for Cyclist Experience Mapping," *ACM Transactions on Sensor Networks*, vol. 6, no. 1, 2009.

[7] K. Shilton, "Four Billion Little Brothers?: Privacy, Mobile Phones, and Ubiquitous Data Collection," *Communications of the ACM*, vol. 52, no. 11, 2009.

[8] P. Mohan, V. Padmanabhan, and R. Ramjee, "Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones," in *Proc. of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys)*, 2008.

[9] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, "Place Lab: Device Positioning using Radio Beacons in the Wild," *Pervasive Computing*, vol. 3468, 2005.

[10] K. L. Huang, S. S. Kanhere, and W. Hu, "Are you Contributing Trustworthy Data?: The Case for a Reputation System in Participatory Sensing," in *Proc. of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWIM)*, 2010.

[11] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma, "Understanding Mobility based on GPS Data," in *Proc. of the 10th International Conference on Ubiquitous Computing*, 2008.

[12] GeoLife GPS Trajectories. [Online]. Available: http://research.microsoft.com/en-us/projects/geolife/

[13] AllJoyn Peer-to-Peer. [Online]. Available: http://developer.qualcomm.com

[14] J. Krumm, "Inference Attacks on Location Tracks," in *Proc. of the 5th IEEE International Conference on Pervasive Computing (Pervasive)*, 2007.

[15] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving Privacy in Participatory Sensing Systems," *Computer Communications*, vol. 33, no. 11, 2010.

[16] A. Beresford and F. Stajano, "Location Privacy in Pervasive Computing," *IEEE Pervasive Computing*, vol. 2, no. 1, 2003.

[17] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: Privacy-preserving Data Aggregation in People-centric Urban Sensing Systems," in *Proc. of the 29th IEEE International Conference on Computer Communications (INFOCOM)*, 2010.