A Context-Aware Privacy Scheme for Crisis Situations

Mohannad A. Alswailim*, Hossam S. Hassanein and Mohammad Zulkernine
School of Computing
Queen's University
Kingston, ON, Canada K7L 2N8
{mohannad | hossam | mzulker} @cs.queensu.ca

***Abstract*— Participatory sensing allows individuals and groups to contribute to an application using their handheld sensor devices. Data collected from participants including their location, time, contacts, etc. are vital to the accuracy of the application but are considered private to the participants. The design of a successful participatory sensing application must consider the challenge of the accuracy-privacy trade-off. In more critical situations when a crisis occurs, however, the accuracy-privacy trade-off becomes more complex. When a participant is at risk, data accuracy becomes more important than participant's privacy. In this paper, we propose a Context-Aware Privacy (CAP) scheme. CAP aims to provide privacy-preserved data to authorized recipients based on the status of participants. Depending on the recipient category, their role and policies enforced, a different level of participants' private data may be received. Experimental results show that the CAP scheme achieves a high level of privacy protection in safe areas. In risk areas/situations the scheme achieves a higher level of data accuracy than existing privacy schemes.**

***Index Terms*— Privacy; differential privacy; crisis response system; participatory sensing.**

I. INTRODUCTION

In recent years, devices such as smartphones and tablets are increasingly being equipped with various embedded sensors such as camera, microphone, GPS, proximity, accelerometer, temperature and humidity [1]. These sensors enable a wide range of applications in Participatory Sensing (PS). PS allows the users of these devices to participate by sensing and collecting data from their surrounding environment and sending them to the application server.

Participants' contributions enhance the PS application services for the end users. These contributions become crucial when the application collects data about a crisis. A crisis such as fires, earthquakes and floods needs an urgent action to be taken to disturb its difficulties. A crisis PS application can be a major source of information for a Crisis Response System (CRS). CRS consists of a group of authorities who are trained to deal with such situations [2]. In addition to the basic pre-exist data at CRS, authorities need data, through the PS application, that is directly related to the crisis and individuals who are within close proximity to the crisis to make rescue plans.

Data collected from participants including location, time, contacts, etc. are significant to the CRS and are considered private to the participants. Protecting participants' privacy, on the one hand, is essential to encourage them to contribute in such applications. On the other hand, data accuracy is vital to execute CRS optimal performance. Therefore, balancing the privacy-accuracy trade-off is challenging especially that participants may become at risk and loos their lives.

To overcome this challenge, we propose a Context-Aware Privacy (CAP) scheme. CAP aims to provide privacy-preserved data to authorized recipients based on the status of participants. Different recipient categories receive a different level of participants' private data.

CAP consists of two major components: (1) context-aware scheme and (2) privacy scheme. The context-aware scheme decides what and how much private data to release to recipients. The privacy scheme protects participants' private data to a certain level based on the context-aware scheme decisions. It applies a manipulated Differential Privacy (DP) function [3] before sending the data to recipients.

CAP is a viable CRS solution for various environmental conditions such as fire disaster, radiation measurement and air quality. As well, it is compatible with multiple types of private data and recipient categories.

We implement the proposed CAP scheme on a fire crisis dataset. Participants, who are around the crisis, sense the air temperature using their sensor devices from different locations and epochs (time periods). Afterwards, participants send the collected data including the metadata such as the participant's location, time and contacts to CAP. It, then, applies its functions to protect participants' privacy and sends its output to the application server. Recipients, including CRS, use the available output to measure the severity of the disaster and make an efficient rescue plan.

We perform experimental evaluations to assess the success of the proposed CAP scheme in controlling the privacy-accuracy trade-off. The results show that CAP scheme achieves a high level of privacy protection in safe areas. In risk areas/situations the scheme achieves a higher level of data accuracy than existing privacy schemes.

The remainder of this paper is organized as follows. In Section II, we discuss several related works. Section III details our proposed context-aware privacy scheme and its related algorithms. In Section IV, we describe the experimental evaluation, setup and the evaluation results. Section V concludes our work.

* Mohannad A. Alswailim is also affiliated with Qassim University (QU), Qassim, Saudi Arabia

978-1-5386-4727-1/18/$31.00 ©2018 IEEE

II. Related work

Protecting participants' privacy is important to encourage them to contribute to PS applications. Participant data can be privacy-preserved before being published to the end users. Due to the lack of researches discussing privacy schemes in crises, we, in this section, discuss the related work that is proposing privacy schemes including DP in various PS applications.

Participant trajectory and position data in PS applications have been privacy-preserved by DP [4, 5]. Li et al. [4] proposed a differentially private trajectory data publishing scheme to protect the privacy of sensitive areas. The scheme is based on partition-based models to partition the original location universe at each time point into multiple groups. The scheme follows an algorithm to select optimal partitions to apply the DP to protect the trajectory privacy. To et al. [5] proposed a scheme to protect the privacy of participant locations in PS applications. They assumed that a trusted third party has access to data sanitized by DP. Thus, the trusted third party can release participant locations to the PS applications in noisy form, according to DP.

Jin et al. [6] proposed a differentially private incentive scheme that preserves the privacy of each participant's bid against others including curious participants within the same application. Some PS applications offer a reward to the participant to do a required task as an encouraging step. To win the reward, participants submit their bids that contain some private and sensitive data to be protected by DP.

Chen and Ma [7] proposed a privacy-preserving aggregation scheme to limit PS applications of learning participants' sensitive data. The scheme applies the concept of DP by adding noise to the sensitive data. Then, it encrypts the noised data and sends them to the application server. The application server can only learn the sum of the noised data.

Existing schemes work on satisfying privacy-preserving PS applications. These schemes do not consider crisis situations on their journey to protect participant privacy. As a result, they do not provide different privacy levels in critical situations when data accuracy is more of a concern.

III. The CAP Scheme

In this section, we overview the proposed CAP scheme in Section III.A. Sections III.B and III.C detail the context-aware scheme and the privacy scheme algorithms, respectively.

A. The CAP Scheme Overview

CAP is a scheme that works in critical situations when crises occur. It aims to balance the privacy-accuracy trade-off challenge based on the status of participants. When a participant is at risk, more accuracy and less privacy will be released to recipients, and vice versa. Different recipient categories receive a different level of participants' private data.

Fig. 1 shows the data flow starting from the participants' data collection passing through the CAP scheme then onto the PS application server. Participants start the process by sensing the required data using their sensor devices. They send the sensed data to CAP to decide what data to release and apply privacy scheme. Then, CAP releases the privacy-preserved data to the PS application server.

In this work, we protect privacy based on both policy and technology protections. Policy protection is by enforcing rules through "rules entity" in the context-aware scheme. Technology protection is by enforcing the privacy scheme on the data that is considered private. Thus, the course of the CAP scheme goes through two major components: (1) context-aware scheme and (2) privacy scheme.

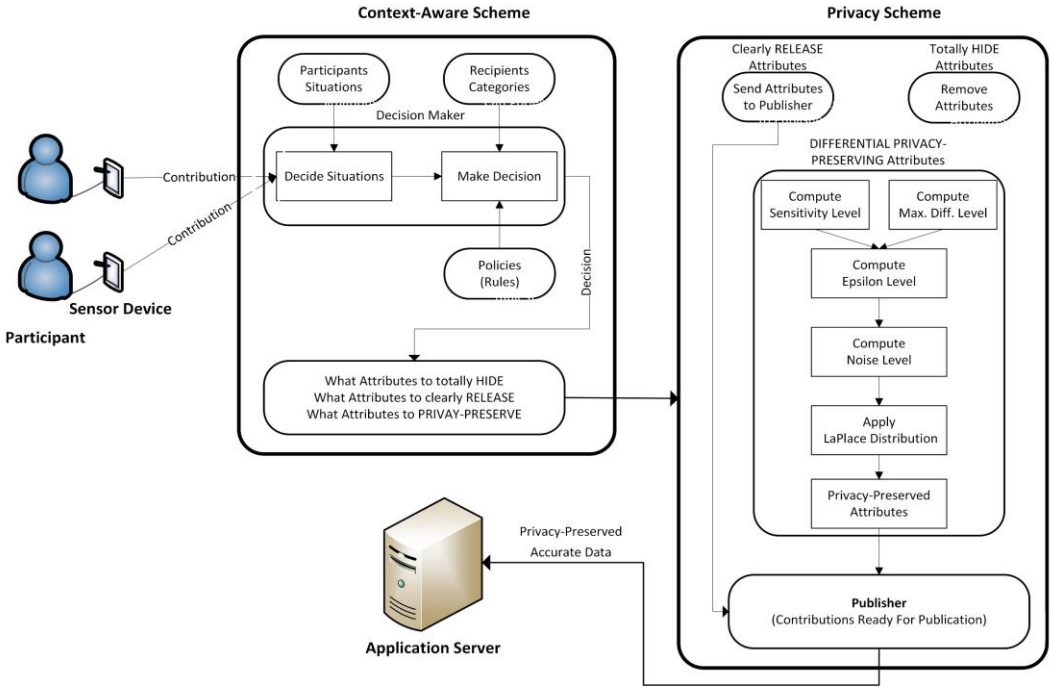


Fig. 1: The CAP architecture

Algorithm 1 Context-Aware Scheme

Input: Participant Contributions
Output: Decisions of what to release to who

1. **Get** Participant Situations
2. **Get** Recipient Categories
3. **Get** Administrator Policies
4. **Get** *P* Attributes

Decision Maker Entity

5. **for** $i \leftarrow 1$ **to** m **do**
6. **for** $j \leftarrow 1$ **to** n **do**
7. **if** $P_i.location \in S_j$ (radius R_j)
8. **then** add P_i to S_j
9. **for** $j \leftarrow 1$ **to** n **do**
10. **apply** $Rule_j$ to S_j
11. **Decide** "what (participants') attributes to release to who (recipients)"

In the first component, the context-aware scheme decides what participant's data to release to who and to what level of privacy protection, as discussed in Section III.B. The context-aware scheme inputs are the participants contributions data and their metadata. Its decision depends on multiple contexts, i.e., participant situations, recipients of the data and a set of policies.

The second component performs three steps on participant data based on the decisions of the first component. Step one removes some data that the first component decides to hide from recipients. This step prevents the selected recipients from accessing that specific participant data. Step two publishes certain data that the first component decides to release in its original format. This step allows the selected recipients to receive those certain data clearly. Lastly, step three applies the privacy scheme on the data that are selected to be privacy-preserved before they reach recipients, as explained in Section III.C.

B. The Context-Aware Scheme

In Algorithm 1, participant situations, recipient categories and policies are defined based on the application requirements. In every epoch**,** the context-aware scheme receives contributions from participants in an affected area. Participants' contributions consist of multiple attributes describing their surrounding environment and themselves.

The context-aware scheme divides the crisis area map into multiple sectors, as in Fig. 2. It centers the crisis location, R_0, and creates n nested circle zones by considering radius R_j, where $j = \{1, \ldots, n\}$. Then, the system splits the map into four cardinal directions (north, east, south and west) to ease following crisis direction and locating participants. We follow the sectors division method in our previous work [8]. As a result, the application acquires n situations (S_j). The closer the sector to the crisis, the higher the risk.

The decision maker entity, within the context-aware scheme in Fig. 1, determines participants' situations based on their proximity to the crisis that can be derived from their location data. The entity forms situation groups by adding a $participant_i$ (P_i), where $i = \{1, \ldots, m\}$, to its relevant S_j. Then, the decision maker entity applies the predefined rules ($Rule_j$) onto S_j. Each of these rules considers participant data attributes, participant situations and recipient categories. For example, the *j-th* set of rules applies to the *j-th* situation that decides what participant attributes to hide, release or privacy-preserve before sending them to recipients. Each recipient category may receive different types of participant attributes and different level of privacy.

In the end, the context-aware scheme formulates three output decisions of participant data attributes, i.e., totally hide, clearly release and privacy-preserve. It forwards these decisions to the privacy scheme component.

C. The Privacy Scheme

The privacy scheme component receives the context-aware scheme output decisions to apply them. It applies the three steps of hiding, releasing and privacy-preserving to each situation at a time. Then, it treats $recipient_k$ (C_k), where $k = \{1, \ldots, d\}$, by checking each participant attribute ($P_i.A$) decision, as in Algorithm 2. If $P_i.A$ decision is to hide, remove the attribute from the published list. If $P_i.A$ decision is to release, forward the attribute to the publisher. If $P_i.A$ decision is to privacy-preserve, apply a manipulated DP function on the attribute, then send the privacy-preserved attribute to the publisher. As a result, the scheme sends the publishable attributes to C_k.

In the case of $P_i.A$ requires privacy preservation, the privacy scheme applies a manipulated DP function. DP is a concept for dataset privacy that learns as much as possible about a group of participants while learning as little as possible about individuals [3].

$$Pr[f(D) \in S] \leq e^{\varepsilon}.Pr[f(D') \in S] \quad (1)$$

A randomized function f gives ε-differential privacy if for all adjacent datasets D and D' differing on at most one participant, and all events $S \subseteq \text{Range}(f)$ [3].

DP requires computing multiple factors, i.e., sensitivity level, maximum difference in an attribute and privacy level, to compute the noise level properly.

One of the factors is sensitivity level (Δf). Sensitivity level is the maximum amount of all possible datasets by which the present or absent of a participant can change the outcome [3].

$$\Delta f = max_{D,D'} |f(D) - f(D')| \quad (2)$$

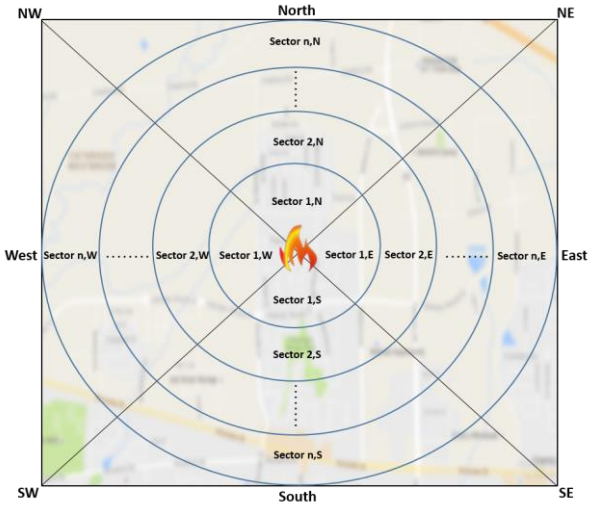


Fig. 2: A crisis map after splitting the area into sectors in each direction

Another factor is the maximum difference between two participants' element (E) of the same attribute (α).

$$\alpha = max_{1\le i,l\le m} |(E_i) - (E_l)|, \quad i \neq l \qquad (3)$$

The privacy scheme cannot use the same noise level for all attributes due to the differences in the nature of the attribute types. Therefore, the purpose of computing α is to relate the noise level to the attribute range values when applying the Laplace distribution.

Privacy level (ε_i), a.k.a. privacy budget, is a key factor in computing how much noise the scheme needs to add to the result to protect participant data privacy. The privacy level is not an absolute measure of privacy but is rather a relative measure. The scheme computes ε_i as the natural logarithm of the ratio of Eq. 1. It is the inverse of the exponential function e^{ε} divided by α.

$$\varepsilon_i = \left| ln \frac{m * \beta_i * \gamma_j}{1-\beta_i} \right| / \alpha \qquad (4)$$

where m is the total number of participants in a situation S_j in one epoch. β_i is the probability of identifying participant P_i in the current contribution. γ_j is a situation degree of danger. γ_j is an assigned value set by the administrator. The purpose of γ_j is to adjust the privacy level to match the situation critical condition that differs from one to another. Thus, each situation has a same value of γ.

By computing the sensitivity level (Eq. 2) and privacy level (Eq. 4), the scheme computes the noise level (b_i). It is a real number that will be added to the true participant data to achieve privacy.

Algorithm 2 Privacy Scheme

Input: Context-Aware Scheme Decision
Output: Privacy_Preserved Data

1. **for** $j \leftarrow 1$ **to** n **do**
2. **for** $k \leftarrow 1$ **to** d **do**
3. **for** $i \leftarrow 1$ **to** $|S_j|$ **do**
4. **if** $P_i.A$ = *Hide*
5. **then** Remove $P_i.A$ from list
6. **else**
7. **if** $P_i.A$ = *Release*
8. **then** Send $P_i.A$ to *Publisher*
9. **else** Apply *DP* on $P_i.A$ and **Call** *Differential Privacy*
10. Send $PP.P_i.A$ to *Publisher* *//PP: Privacy-Preserved*
11. **Send** P_i attributes in *Publisher* to C_k

Manipulated *Differential Privacy* Function

Compute sensitivity level (Δf)
12. $\boldsymbol{\Delta f = max_{D,D'} |f(D) - f(D')|}$

Compute maximum difference between two elements (E) of the same attribute (α)
13. $\boldsymbol{\alpha = max_{1\le i,l\le m} |(E_i) - (E_l)|}$

Compute privacy level (ε_i)
14. $\boldsymbol{\varepsilon_i = \left| ln \frac{m * \beta_i * \gamma_j}{1-\beta_i} \right| / \alpha}$

Compute noise level (b)
15. $\boldsymbol{b_i = \frac{\Delta f}{\varepsilon_i}}$

Compute *LaPlace* mechanism
16. $\boldsymbol{Lap(\mu, b_i)}$

$$b_i = \frac{\Delta f}{\varepsilon_i} \qquad (5)$$

To add b_i to the original data, the scheme applies *Laplace* distribution. It is mainly how wide is the noise to be added to protect privacy.

$$Lap = (\mu, b_i) \qquad (6)$$

where μ is the position that depends on the original data that needs to be privacy-preserved. b_i is the scale factor that depends on Δf and ε_i, not on the dataset. The higher the b_i, the flatter the scale is, and the lower the b_i, the sharper the scale is.

IV. Experimental Evaluation

In this section, we discuss our experimental evaluation of the CAP scheme. Section IV.A reviews the implementation setup and evaluation environment. In Section IV.B, we describe the dataset we use in the implementation. Finally, we discuss the implementation results and the evaluation metrics to assess the success of our scheme in Section IV.C.

A. Evaluation Environment

Our implementation of the CAP scheme uses a dataset by a group of participants who are within the vicinity of a crisis (see Section IV.B for details). The participant contributes by collecting sensor data, location, date and time. In each participant contribution, metadata, i.e., age, gender, height, weight and health condition are included. Participants send their contributions to CAP periodically. Then, CAP applies the steps of the context-aware scheme and the privacy scheme as discussed earlier. Based on the CAP decisions of what data to release and to what level of privacy protection, the publisher component sends the privacy-preserved data to the application server.

B. Dataset

Our overall dataset consists of two parts, sensors dataset and participants dataset. Sensors dataset considers the crisis environment and sensor data, i.e., location, date and time. Participants dataset deliberates participants personal profiles, i.e., age, gender, height, weight and health condition. Our dataset is publicly available on Scholars Portal Dataverse[1] [9]

i. Sensors Dataset

Due to the lack of data collected from sensor devices of an actual crisis, we generated our dataset following some of the early steps of our previous work [8]. In our generated dataset, we randomly assigned the crisis location and the participants located within the vicinity of the crisis.

Our use case is a fire crisis. We generated a heat map by creating three heat levels in a radius of 2 km from the crisis. The dataset contains 250 participants contributing for six days. The data collector gathers the data every 2 minutes (epoch). In every epoch, participants use their smartphones' temperature sensors to collect sensor data in addition to their location, date and time. In our use case, the temperature is an instance. Participants freely travel from one position to another. Therefore, the number of participants in any given zone differs from one epoch to another due to the absence of some of the participants who are out of the crisis considered range.

1 https://doi.org/10.5683/SP2/DIDFP9

We apply a Gaussian distribution to generate a temperature value for every participant in every epoch. In the Gaussian distribution, we assign a mean (μ) that corresponds to the ground truth temperature for every heat level and a standard deviation (δ) that corresponds to the limited possible error of the participant contribution.

ii. Participants Dataset

Metadata, in this work, is the data about participants that do not change frequently. To have such data, we searched for a real-world dataset that contains data about actual people collected by official organizations. Statistics Canada has collected a dataset "Canadian Community Health Survey" (CCHS) in 2014 and published it in 2016 [10]. We selected specific attributes to describe the participants. These are age, gender, height, weight and health condition. We selected 250 out of thousands of participants in the survey and assigned their data randomly to our 250 participants. The participants dataset eventually has one record to each participant.

To make our dataset more challenging to privacy, we applied two steps that make some participant contributions distinct from others that may cause, with some effort, to re-identify the participants. In the first step, we chose 80 uncommon characteristic participants in the survey (among the selected 250) that are either underweight or obese. We computed the Body Mass Index (BMI) values to obtain the two uncommon characteristics. In the second step, for each participant, we assigned a visit probability to each sector in the map. We allocated random adjacent sectors to each participant and random probability to each sector. Participants, selected by these two steps, are more vulnerable to be re-identified than others due to their uncommon characteristics and appearance probability in some sectors.

C. Experiment Results

At the end of the CAP process, every publishable data will be sent in either its original or privacy-preserved format. Therefore, we need to consider the consequences that those publishable data may cause regarding protecting participant privacy and rescuing victims. Hence, we evaluate the success of our scheme by measuring: (1) the possibility of re-identifying participants due to releasing some of their data in original format, and (2) the impact that the privacy scheme may cause to prevent rescuing victims due to hiding full or part of the original data.

Identification Confidence (*IC*) is a metric to measure the confidence level of re-identifying a participant through its published data.

To apply IC, we need to classify all the participant attributes and measure how sensitive each attribute is in the two possible release formats, original and privacy-preserved. As a result, we understand the *quasi-identifier* attributes that may lead to re-identify participants. A quasi-identifier attribute is an attribute that can be used to probabilistically identify a participant either by itself or in combination with other attributes. Table 1 shows all the attributes in our use case and their sensitivity parameters in the two release formats. The proposed sensitivity parameters are high, medium and low. For example, if combining a high sensitive attribute with another high sensitive attribute, the result can lead to a high probability of re-identifying the participant.

Data Format / Attributes	Original	Privacy-preserving
Participant ID	Hi	Med
Sensed Data	Low	Low
Contact	Hi	Low
Location	Hi	Low
Time	Hi	Low
Date	Hi	Low
Age	Hi	Low
Gender	Hi	Med
Weight	Med	Low
Height	Med	Low
Health Condition	Med	Low

Table 1: Sensitivity parameters attributes in the two release formats

Sweeney et al. [11] found that the probability of re-identifying an individual by combining the quasi-identifiers of the date of birth, gender and full postal code is 87%. By making either the date of birth or full postal code less specific, the probability drops to 44%. Hence, decreasing the number and/or the details of the quasi-identifiers will result in decreasing the probability of re-identifying individuals. Thus, we assign the values of 80, 40 and 10 to the attribute parameters high, medium and low, respectively. These values allow the IC metric to compute the overall sensitivity of combining multiple attributes.

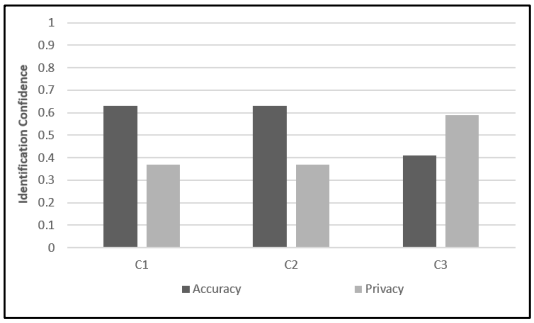
We set two thresholds, 30% and 60%, to decide the probability of re-identifying participants. In this case, a small value means good privacy and poor accuracy level, and vice versa. If the IC value is below 30%, then the probability of re-identifying a participant is low, that privacy protection is good. If the value is between 30% and 60%, the probability to re-identify a participant is medium, that the privacy and accuracy are fair. Finally, if the value is above 60%, the probability to re-identify a participant is high that the accuracy is good.

To compute IC metric, we take the average parameter values of all publishable attributes for each recipient category in every situation, as in Eq. 7:

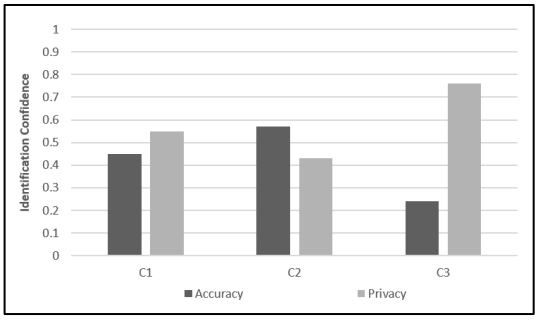
$$IC_{j_k} = \frac{\sum publishable\ attribute\ values}{total\ number\ of\ publishable\ attributes} \quad (7)$$

Where $j = \{1, 2, 3\}$ referring to the number of situations that are S_1: high risk situation, S_2: moderate risk situation and S_3: safe situation. Where $k = \{1, 2, 3\}$ referring to the number of recipient categories that are C_1: family and friends, C_2: authorities and C_3: journalists and public. Authorities in C_2 are the most important recipient category because they are responsible for setting rescue plans regardless of the participant situations. Therefore, the application administrator allows more data to be released to this category than others. Family and friends in C_1 come next in releasing data due to the close relationship, then journalists and the public in C_3.

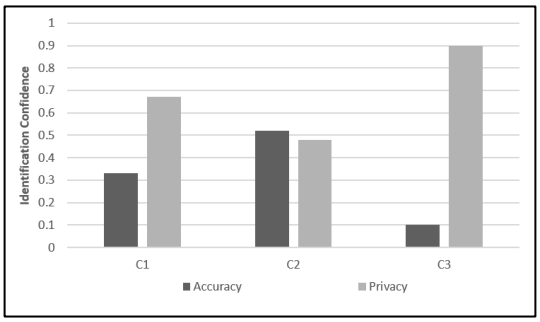
As a result, IC values in S_1 show that C_1 and C_2 receive more accurate data than C_3 and the probability of re-identification is high at 63%, 63% and medium at 41%,



a. High Risk Situation (S_1)



b. Moderate Risk Situation (S_2)



c. Safe Situation (S_3)

Fig. 3: Privacy-accuracy trade-off in three situations based on IC values

respectively, as shown in Fig. 3.a. At S_2, the results show that C_2 receives less accurate data than in S_1 because participants status are more comfortable; however, it receives more data than other categories in the same situation. The IC values are medium at 45%, 57% and low at 24% for C_1, C_2 and C_3, respectively, as in Fig. 3.b. Finally, S_3 is the most comfortable and safe situation for participants, and the results show that C_2 receives more accurate data than others to measure the crisis severity even if the participants are safe. The IC values are the lowest, as shown in Fig. 3.c, and the privacy level is good due to low risk on participants.

From the IC metric results, we can notice that the impact the privacy scheme may cause in preventing rescuing victims is very low especially in S_1, the most critical situation. The probability of re-identifying participant is as high as 63% for C_1 and C_2, which means that the published data is critically managed to provide more accurate data besides privacy. In the other two situations, S_2 and S_3, risk recedes to a level that rescuing participants becomes less important. Thus, the IC values drop to the range of 57% and 10% to all recipients.

V. Conclusion

Protecting participants' privacy is essential to encourage them to contribute in crisis PS applications. Also, data accuracy is necessary to set finest plans by rescue personnel. Therefore, balancing the privacy-accuracy trade-off is challenging especially that participants may become at risk in some situations. We proposed the CAP scheme that aims to provide privacy-preserved data to authorized recipients based on the status of participants. Different recipient categories receive a different level of participants' private data. We implemented the proposed CAP scheme on a fire crisis dataset. Experimental results showed that the CAP scheme achieves a high level of privacy protection in safe situations, and a higher level of data accuracy than existing privacy schemes in risk situations.

Acknowledgment

This research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number: STPGP 479248. The findings achieved herein are solely the responsibility of the authors.

References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory Sensing," in *Proceedings of the ACM International Workshop on World-Sensor-Web*, pp. 117–134, 2006.

[2] J. E. Hale, R. E. Dulek, and D. P. Hale, "Crisis response communication challenges: Building theory from qualitative data," *The Journal of Business Communication*, vol. 42, no. 2, pp. 112–134, 2005.

[3] C. Dwork, "Differential privacy," in *Proceedings 33rd International Colloquium Automata, Languages and Programming, ICALP*, pp. 1–12, Springer Berlin Heidelberg, 2006.

[4] M. Li, L. Zhu, Z. Zhang, and R. Xu, "Differentially private publication scheme for trajectory data," in *IEEE First International Conference on Data Science in Cyberspace (DSC)*, pp. 596–601, June 2016.

[5] H. To, G. Ghinita, and C. Shahabi, "A framework for protecting worker location privacy in spatial crowdsourcing," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.

[6] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pp. 344–353, 2016.

[7] J. Chen and H. Ma, "Privacy-preserving aggregation for participatory sensing with efficient group management," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 2757–2762, Dec 2014.

[8] M. A. Alswailim, H. S. Hassanein, and M. Zulkernine, "A participant contribution trust scheme for crisis response systems," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2017.

[9] M. A. Alswailim, H. S. Hassanein, and M. Zulkernine, "Sensor data and participants profiles," in *Scholars Portal Dataverse*, 2018.

[10] "Statistics canada. 2016. canadian community health survey, 2014: Annual component [public-use microdata file]. ottawa, ontario: Statistics canada. health statistics division, data liberation initiative [producer and distributor]."

[11] L. Sweeney, A. Abu, and J. Winn, "Identifying participants in the personal genome project by name," *In Harvard University, Data Privacy Lab*, 2013.