A Survey on Participatory Sensing Systems

Asha Raj[1], Abeera V P[2]

[1]M.Tech Student, KMEA, Computer Science and Engineering, Mahatma Gandhi University, Kerala, India

[2]Assistant Professor, KMEA, Computer Science and Engineering, Mahatma Gandhi University, Kerala, India

Abstract: ***Wireless networks consist of a collection of large number of sensor nodes. The emergence of various sensors made the emergence of participatory sensing systems. Sensor data, in its original form, contains sensitive information about individuals. Privacy protection is very important for participatory sensing systems. In wireless networks, the nodes communicate with each other through wireless medium. Data aggregation helps in reducing the number of bits transmitted thereby reduces the total energy consumption. In this survey, we summarize different privacy preserving techniques and data aggregation protocols for wireless sensor networks. We also provide a brief description of Reed-Solomon erasure coding technique to detect and correct errors in transmission.***

Keywords: Participatory sensing, privacy preserving, collaborative path hiding, data aggregation, erasure coding

1. Introduction

Participatory sensing is an emerging paradigm where group of people contributes sensory information. With the growth in mobile devices, such as smart phones, which has multiple sensors, the demand for participatory sensing has increased. Participatory sensing systems consist of multiple mobile users gathering data in a joint way. Participatory sensing has been widely used in many applications such as health, traffic, noise and weather monitoring, community service, and many other applications. Two most important challenges with participatory sensing systems are:

- Privacy and quality preservation
- Variety of sensing data

Sensing record consists of data along with spatial and temporal information. If the service provider is not honest, he may infer the private information of the user participating in participatory sensing applications from these location and time information. Because of this, many users are unwilling to contribute data for participatory sensing systems. But quality of service can be guaranteed only if there are enough number of participants. Therefore, privacy preserving is very important in participatory sensing. Variety of sensing data is another major challenge. Sensing data may include temperature, location, time, digital images, videos, etc. This paper describes various privacy preserving techniques, collaborative path hiding techniques and different data aggregation protocols for participatory sensing systems. Finally, it also describes Reed-Solomon coding for detecting and correcting errors.

2. Privacy Preserving Techniques

A number of privacy preserving techniques have been proposed to address the privacy and quality of sensing data for participatory sensing systems. These techniques can be classified as follows:

2.1 Randomization Technique

K. Mivule [1] proposed a noise addition technique for data privacy. This method is also called as Data Perturbation technique where the data will be modified so that it no longer represents the real world. It is also known as noise based technique where noise will be added to the original data so that the values cannot be guessed from the distorted data. Figure 1 shows a general data privacy method which can be achieved in 2 steps:

- Data De-Identification
- Noise Addition

Data De-Identification is the process of removal of sensitive information such as personal identification information from the original data. In-order to ensure higher level of confidentiality, noise addition is also introduced. It is the process of adding or multiplying a randomized number to confidential quantitative attributes so that the original data cannot be guessed from the deformed data. For example, if the age attribute is 30, randomly adding a value of 50 with it converts the value to 80. One of the major disadvantages of this method is that original data cannot be reconstructed. Some of the large data collection organizations such as Census Bureau omit sensitive information by using this technique before releasing their statistics to the public.

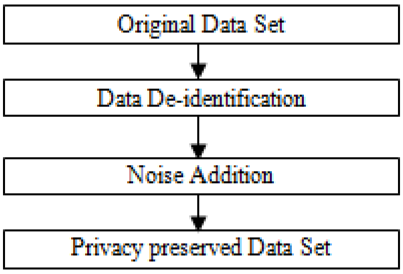


Figure 1: Randomization Technique

2.2 Generalization Technique

L. Sweeney [2] proposed a k-anonymity model for protecting privacy. This method is also called as Anonymization technique. Generalization technique is the act of converting a

http://dx.doi.org/10.21275/v5i6.NOV164370

value from a finer granularity to a coarser equivalent. Converting a street level location value to a city level equivalent is an example of generalization. This technique is applied to participatory sensing system to implement k-anonymity. By k-anonymity it means that it is difficult to distinguish each record from k-1 other records. Table 1 shows an example of 3-anonymous report. Here, the 'Time' values are anonymized to get the 'Generalized time'. For example, 10:30 is represented by the time interval 10:00 – 11:00. Before sending this report to any application, the real value of time will be removed from the report. One of the major disadvantages of this method is the need for an honest third party anonymizer for performing the anonymization technique, which is not always possible in case of a semi-honest model.

Table 1: Generalization Technique

User Id	*Time*	*Generalized Time*	*Tile Id*
1	10:01	(10:00-11:00)	1
2	10:30	(10:00-11:00)	1
3	10:55	(10:00-11:00)	1
4	22:15	(22:00-23:00)	2
5	22:40	(22:00-23:00)	2
6	22:45	(22:00-23:00)	2

2.3 Cloaking Technique

Xu, Ge [3] proposed a location cloaking method for protecting location privacy in the context of Location-Based services (LBS). Due to the emergence of smart phones, LBS have become one of the most popular mobile applications. When user requests a service, the location details are also captured. For example, when user clicks a photo using his smart phone camera, the time, date and location where the photo was clicked will also be automatically embedded in the photo. Cloaking technique replaces the actual location value with a larger area. Users can also configure their mobile devices as to when and to whom the location information should be published. One of the major disadvantages of this method is that even though the privacy is protected, the quality of the reported data is reduced.

2.4 Cryptographic Technique

Rastogi et al. [4] proposed cryptographic method for privacy preservation. End-to-end encryption can provide high security of reported data. Before sending the report, at the sender's side, the report is encrypted. At the receiver's side, the report is decrypted. Cryptography protects the content of the report from being disclosed to any unauthorized entity. It ensures data integrity, accuracy and confidentiality. One of the major disadvantages of this method is that it protects data only from external attacks, such as eavesdropper attack and does not protect from internal attacks, such as service provider attack and participants' attack. Thus it fails to prevent service provider and other participants' from inferring users' sensitive data.

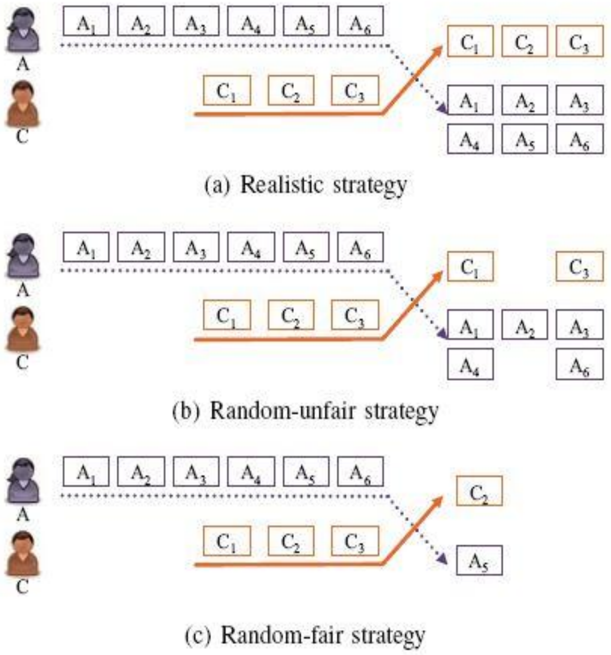
3. Collaborative Path Hiding Techniques

Christin et al. [5] proposed various exchanging strategies and reporting strategies for protecting the location privacy of the participants. Exchanging strategies deal with different ways of exchanging the sensor readings between the participants. Reporting strategies deal with different ways of reporting the sensor readings to the server.

3.1 Exchanging Strategies

In this technique, the participants collaborate to protect their privacy. Figure 2 shows different exchanging strategies. It uses the concept of path jumbling where location privacy is preserved in a decentralized way by exchanging the readings between the participants. Thus it breaks the connection between the spatiotemporal information and the identity of the user. Spatiotemporal information indicates the time and location information at which the sensor readings were taken. Different strategies to exchange the sensor readings to the application are as follows:

- Realistic Exchange Strategy: In this method, participants exchange their entire set of collected sensor readings at each meeting.
- Random-unfair Exchange Strategy: In this method, each participant randomly determines the number of reports he wants to exchange. Each participant may exchange different number of reports.
- Random-fair Exchange Strategy: In this method, the participants agree on a common number of n reports to exchange at each meeting. Here, the two participant exchange equal number of reports.



(a) Realistic strategy

(b) Random-unfair strategy

(c) Random-fair strategy

Figure 2: Exchanging Strategies

3.2 Reporting Strategies

Different ways of reporting the sensor readings to the server are as follows:

- Time Based Strategy: In this method, the sensor readings are periodically (hourly/daily) reported to the server. It ensures that the application receives readings on a timely manner. One of the major drawbacks is that, during the time period, if jumbling did not happen, there is a chance that the sensor readings could not be exchanged with other participants, thereby reaching the server directly from the participant itself.

- Exchange Based Strategy: In this method, the sensor reports are reported to the server after every meeting. It is also known as 1-Exchange strategy. This strategy ensures that the reports are jumbled before it reaches the server. One of the major drawbacks is that, if the meeting is delayed, it could result in long reporting latency.
- Metric Based Strategy: In this method, the reports are reported to the server after a particular threshold value is reached. For example, if the percentage of jumbled reports reaches a given threshold (Jumbling based), or the distance between each location and the jumbled path is above a threshold value (Distance based).

4. Data Aggregation Protocols

Data aggregation protocols can be divided into two:
1. Tree based data aggregation protocols
2. Cluster based data aggregation protocols

Tree based data aggregation protocols consists of parent nodes and leaf nodes. Here, data aggregation is being performed by intermediate nodes. Cluster based data aggregation protocols consists of different clusters. Data aggregation is performed locally at each cluster.

Patel et al. [6] proposed data aggregation techniques which deal with collecting and aggregating data. It is been widely used in wireless sensor networks. Security is one of the major concerns of data aggregation. Cryptographic techniques are used to achieve security. Some of the protocols that provide security along with data aggregation are listed below. These protocols are designed for static networks and are not suitable for participatory sensing where network changes dynamically.

4.1 Hop-to-Hop secure data aggregation protocols

In this method, data encryption and decryption is done between each pair of nodes in the network. It implements a key based mechanism (Pair-wise keying) which ensures data confidentiality. As the intermediate nodes have to decrypt the data, it offers more chances to the attackers to get the sensor data.

4.2 End-to-End secure data aggregation protocols

This method is more flexible than hop to hop data aggregation method. Once sensor data is encrypted at the sender side, it is decrypted only at the service provider. End to end data privacy is achieved through Homomorphic encryption. It allows performing arithmetic operation on encrypted data without the need for decrypting it. As it does not require the intermediate nodes to decrypt the data, it is more secure compared to hop to hop data aggregation method.

5. Erasure Coding

I. Reed and G. Solomon [7] proposed Erasure coding method for participatory sensing systems. It breaks a sensing record into fragments and encodes with redundant data pieces. A stream of data in the form of 0's and 1's are transmitted over a communication channel. Errors can occur in the transmitting channel causing the bits to change. ie; converting the 0's to 1's and vice versa. In order to check whether the original data has been changed, redundancy has been introduced. It helps in recovering the original data in case of error. Each bit is sent 'n' times in sequence, and the bit that occur the majority of the time is selected. For example, if a bit is sent 3 times with values 0, 1 and 0, then the actual bit is considered as 0 as it occurred twice out of 3 trials. Figure 3 shows the flow diagram for encoding and decoding technique.

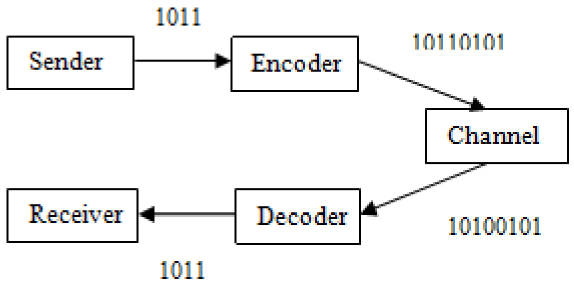


Figure 3: Erasure Coding

The sender sends the source data which is then passed through the encoder. Encoder encodes the source message into codeword, which adds redundancy in order to detect and correct errors in transmission. When the data is passed through the channel, it may introduce several errors. Decoder corrects the errors and reclaims the source message. The original data can be decoded from any k out of m encoded slices, where k is approximately equal to the size of the original record and m is the number of redundant data pieces and m>k. Finally, the receiver receives the original source message. One of the main features of Reed-Solomon codes is that, here redundancy occurs naturally.

6. Conclusion

In this paper, various privacy preserving techniques, collaborative path hiding techniques and data aggregation protocols are presented. Pros and cons of randomization, generalization, clocking and cryptographic techniques are discussed. We have seen that data perturbation technique with noise addition is used to provide privacy for data sets. In order to protect the location privacy of users who contribute to participatory sensing systems, a collaborative and decentralized approach is used. Depending on the nature of the application, different privacy preserving techniques are adopted. Sensor readings are exchanged between participants in order to mask their paths. Various exchanging strategies and reporting strategies are also presented here. Among all the reporting strategies, threshold based approach provides strong protection of the sensor readings. Depending on the privacy needs of the application and the degree of trust in other participants, the exchanging and reporting strategies are determined. Various secure data aggregation protocols are also discussed here. Based on the requirements, the desired protocols can be selected easily.

References

[1] K. Mivule, “Utilizing Noise Addition for Data Privacy, an Overview”, Bowie State University, 14000 Jericho Park Road Bowie, MD 20715, Mivulek0220@students.bowiestate.edu

[2] L. Sweeney, “k-Anonymity: A model for protecting privacy,” Int.J. Uncertainty Fuzziness Know. Based Syst., vol. 10, no. 5, pp. 557–570, 2002.

[3] Xe, Gu, “Location cloaking for location privacy protection and location safety protection”, Iowa State University Digital Repository @ Iowa State University, 2010.

[4] V. Rastogi and S. Nath, “Differentially private aggregation of distributed time-series with transformation and encryption,” presented at the ACM SIGMOD Int. Conf. Manag. Data, IN, USA, Jun. 2010.

[5] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, and S. S. Kanhere, “Privacy-preserving collaborative path hiding for participatory sensing applications,” presented at the 8th IEEE Int. Conf. Mobile Ad-hoc Sen. Syst., Valencia, Spain, Oct. 2011.

[6] K. J. Patel, N. M. Raja,”An Overview of Secure Data Aggregation in Wireless Sensor Network”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015.

[7] I. Reed and G. Solomon, “Polynomial Codes over Certain Finite Fields”, March 31, 2000.

[8] F. Qiu, “Privacy and Quality Preserving Multimedia Data Aggregation for Participatory Sensing Systems”, IEEE Transactions On Mobile Computing, Vol. 14, No. 6, June 2015.

Author Profile

Asha Raj received the Bachelor of Technology degree in Computer Science and Engineering from Cochin University of Science and Technology in 2005 and currently doing Master of Technology in Computer Science and Engineering from Mahatma Gandhi University.

Abeera V P received the Bachelor of Technology degree in Computer Science and Engineering from Government Engineering College, Wayanad in 2007 and Master of Technology in Computer Science and Engineering from Amrita Vishwa Vidyapeetham University, Coimbatore in 2010. She is currently working as Assistant professor in Computer Science Department, KMEA Engineering College, Kerala.