I J C T A, 9(37) 2016, pp. 205-212

A survey on Privacy Preserving Data Aggregation Schemes in People Centric Sensing Systems

K.R. Jansi[*] and S.V. Kasmir Raja[**]

Abstract: The technological advances in mobile and wireless communication have paved way for a new era of large scale sensing network. Smart phones used by people are nowadays equipped with rich set of sensing capabilities. These mobile devices carried by people can be used together to form a network called people centric sensing network. Since people are the mobile custodians, privacy becomes a threat to the network. Privacy protection is important for user's participation which supports large set of cooperative sensing applications. This paper focuses on a study about the various privacy preserving aggregation schemes existing in other domains of the wireless network including people centric sensing. The design goals and challenges that arise when attempting to protect the user's sensitive data are analyzed. The techniques commonly used to achieve user's privacy are the homomorphic property of encryption schemes, Data slicing and Mix schemes. A study about the existing privacy preserving aggregation schemes enables us to design a new method that is more suitable for people centric sensing network. The privacy preserving aggregation scheme should support spatial and temporal aggregation with fault tolerance. The scheme has to be adaptable for both additive and non-additive statistical aggregation functions. Also maintaining the data integrity, data privacy and data accuracy with less computational and communication overhead is essential.

Index Terms: Privacy preserving, aggregation, People centric sensing system.

1. INTRODUCTION

The advance in wireless communication and mobile computing has given rise to several new urban scale applications that benefit the users in his day to day activities. Today's smart phones carried by people has increasing computation and communication facilities equipped with rich set of sensors like GPS, camera, gyroscope, accelerometer, proximity sensor, compass etc. Taking advantage of the facilities of mobile devices, a cooperative network can be formed by the users. People can apply these devices to form a new sensing network called People centric sensing network [3]. Here, the community of people contributes sensor information to infer some knowledge. Such a network offers many new opportunities for cooperative sensing applications [2].

People centric sensing (PCS) has two basic approaches based on the nature of collected data and the degree of user involvement in the sensing process. When users are directly involved in the sensing process, it is referred to as Participatory sensing [1]. When users are not directly involved in the sensing process, it is referred to as opportunistic sensing [1]. There are four main characteristics [2] of People centric sensing. First, PCS uses the existing infrastructure like smart phones and vehicular systems which in turn avoid the deployment cost of the network. Second, Mobility is a driving factor to gain scalability and sensing coverage. The computing, communication and sensing can happen anytime anywhere with very limited multihop wireless communication. Third, the mobile devices are regularly charged and not so energy constrained. Fourth, PCS is application agnostic where security and privacy are the important factors that drive the growth of the network.

[*] Department of Computer science and Engineering, SRM University, Chennai, India. *Email: jansi.k@ktr.srmuniv.ac.in*
[**] Dean, Research, SRM University, Chennai, India.

PCS Vs WSN: The characteristics of PCS are quite different from traditional wireless sensor networks (WSN) [19]. PCS is much suitable for large scale application. In PCS, the system device belongs to individuals with different interests. The system devices can be charged regularly and hence has more powerful energy resource. Sensing data are more related to interactions between people and between people and their surroundings. System node mobility is dynamic and people are not just data users but also active data contributors. In WSN, system devices are owned and managed by a single authority. Here, the sensor devices are not charged regularly and network is mostly static. The sensing data is related to some physical phenomena of interest like temperature, moisture etc. People are only passive users of the data generated by the sensors.

Motivation: Mobile device collects information that is nearly related to user's day to day activities. The context information about the people and their surroundings can be captured from the data generated by the sensors of mobile devices. The inference made from the data creates the concern of user privacy which is a challenge for People centric sensing networks. If the privacy of the users is at risk then the users will not be willing to contribute data. Also users do not have a direct benefit from reporting data. So, the privacy of user has to be guaranteed for the success of the people centric sensing networks.

PCS applications [4] mainly focus on public, private and social aspects of the environment. In many scenarios aggregation statistics has to be periodically calculated from the stream of sensor data published by users to infer a particular pattern or a phenomenon. Example, to plan outdoor activities the average or maximum level of air pollution is used. For traffic management [4], the average speed of vehicles can be used as an indicator of congestion to plan the traffic free route. PCS enables applications [5] such as environmental monitoring, traffic monitoring, health care and so on.

2. MODEL AND DESIGN GOALS

In this section, the basic system model, adversary model and design goals to be satisfied in PCS are discussed.

A. System Model

The procedure of data aggregation [17] in PCS is more likely to be disturbed by the adversary. The basic system model for data aggregation is depicted in Figure 1. Here the users who has the custody of the mobile device are called the mobile nodes (MN).The MN's contribute sensory information to the aggregation server (AS) where the aggregation statistics is computed for the request received from the corresponding service provider. The service provider is the third party entity which processes the request given by the client and directs it to the AS. The Peer-to-Peer communication and communication between AS and MN are possible through WIFI or Bluetooth standards. In PCS, client requests may be handled by various third party entities and aggregation server which creates a threat to the user's sensitive data contributed by the mobile nodes.

B. Threat Model

A people centric sensing system is likely to be attacked by the adversary during the process of data aggregation. In PCS, there are two types of attacks either internal or external attacks. If the adversary is a system entity then it is referred as internal attacks [17]. A node and the aggregation server may be curious, malicious or both. If the adversary is not the system entity, it is referred as external attacks [17]. They might be a third party entity who is interested with the data shared by the community. Eavesdropping with the communication between mobile node and aggregation server is the common attack. Encryption

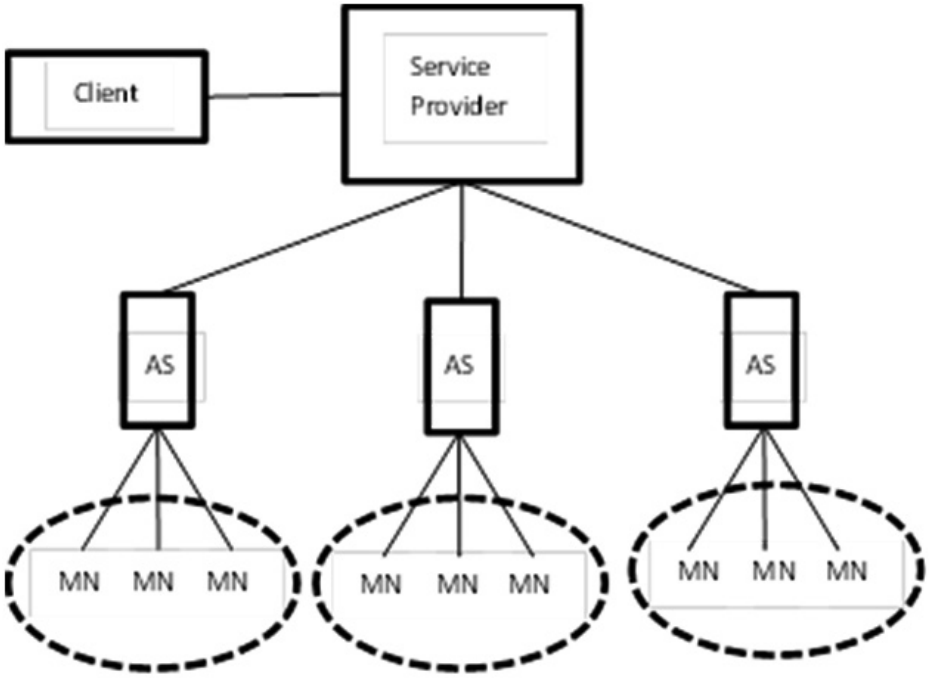

Figure 1: Basic system model

schemes can be used to provide security to the system. False data injection attack, forgery attack can cause damage to the data integrity and data accuracy. Differential attacks [25] have to be considered during the privacy preserving data aggregation. In PCS each user has to contribute the sensor data of their device. The assumption of Trusted Platform Module [17] in the device ensures the integrity of the data generated by the sensors and the users.

C. Design Goal

Any privacy aggregation has some desirable requirements to be fulfilled during the process of data aggregation. The design goals to be considered during the aggregation in PCS are the following

- Aggregation accuracy: The aggregation result computed by the aggregation server should be correct.
- Data Integrity: The privacy preserving scheme should resist false data injection attack and the data contributed by the user and the data received by the AS should be same.
- User Privacy: The user's sensitive data to be confidential and the identity of the user and the data not to be revealed.
- Support Multifunctional data aggregation: The scheme may support additive aggregation functions like sum, average, variance etc. Also schemes need to be extended for non-additive functions like max/min, median, histogram etc.
- Spatial and temporal aggregation: Support for time series data and spatial aggregation is necessary.
- Fault tolerance: Even in the absence or failure of mobile nodes or communication links, the system should be able to proceed with the data aggregation. The system should be reliable and robust.
- Differential privacy: Differential attack is caused by analyzing the aggregated data and tries to reveal the Individual user data. The privacy preserving aggregation scheme should resist differential attacks.
- Dynamic joins and Leaves: Addition and deletion of users in the system should be efficient and dynamic.
- Efficiency can be measured in terms of low communication and computation overhead.

D. Data Aggregation Process

The system is initialized with security parameters at the initial phase. Then client request is received by the service provider. The service provider identifies the corresponding aggregation server and forwards the request to it. Base on the request by the aggregation server the users generate the data and report it to the aggregation server. Aggregation Server computes the statistical aggregation in the privacy preserving manner. The Aggregation Server then forwards the report generated to the service provider.

3. PRIVACY PRESERVING AGGREGATION SCHEMES

This section covers the existing related privacy preserving aggregation schemes in various domains of wireless networks.

A. Smart grid

MUDA [13]: In smart grid communication new solutions for intelligent electricity generation, transmission, distribution and utilization have evolved. The real time power usage pattern of private users is collected in a particular pattern. There is a need for privacy preserving solutions to compute aggregation statistics without disclosing individual user's data. Le Chen et al. have taken advantage of the homomorphic property of Boneh Goh Nissim cryptosystem and bilinear map of Composite order group to provide confidentiality for users reported data. MUDA supports multifunctional data aggregation schemes including average, variance and one-way ANOVA aggregation. The scheme does not support non-additive aggregation statistics. MUDA is also extended to resist differential attacks. MUDA is more efficient in terms of communication overhead and also preserves user's data in smart grid communication.

EPPA [16]: Certain applications use data that are multi-dimensional in nature. Traditionally when multiple dimensions are present each of the dimensional data has to be processed separately. EPPA Scheme adopts the homomorphic Paillier Cryptosystem to achieve the privacy preserving multi-dimensional aggregation for secure smart grid communication. EPPA integrates a multi-dimensional data into one dimensional structure using super increasing sequence when the user data report is generated. Compared with the existing one dimensional schemes used in wireless sensor networks, it reduces the communication and computation overhead. In smart grid, thousands of users may communicate with a single gateway around a residential area. The compressed data aggregation scheme under the public key system is used which considerably reduces the initialization efforts thereby increasing the reliability. According to the security requirements, confidentiality, authentication and data integrity is achieved. To keep the authentication cost minimal, Batch verification is used to verify the validity of the user by the local gateway. The time taken is reduced from 2n to n+1 times for n users. Only external attacks are considered with the assumption of trustable gateway and honest users. EPPA supports only summation operations and also not tolerant to user failures. EPPA cannot be applied to compute multifunctional aggregation and also does not preserve differential privacy. Internal attacks to be analyzed in the system.

PDAFT [12]: Adopts Paillier encryption which is a popular public key encryption scheme. It uses the homomorphic properties of the encryption scheme to encrypt the user data so that the control center can obtain only the aggregated data. PDAFT is fault tolerant to users and control center servers. It is achieved by assuming that the adversary can compromise only minority of the servers. The control center servers are considered as powerful entities which will shed huge cost for an adversary to compromise more servers. Based on the secret sharing method atleast d+1 share is needed for recovering the original data. The consideration of malfunction of both users and servers helps in increasing the reliability and support fault tolerance. PDAFT also supports temporal and spatial summation aggregations against strong adversary. It can also be extended to support dynamic users. As a security requirement, confidentiality is achieved

with communication effectiveness. The transmitting time for the data collection process is significantly reduced when the numbers of users are more. PDAFT does not provide multifunctional aggregation and it does not preserve differential privacy. PDAFT supports only single dimensional and summation aggregation.

B. Wireless Sensor Networks

PDA [19]: In wireless sensor network, preserving privacy during data aggregation is a challenging issue. Wenbo He et al. [19] consider cooperative sensing applications where privacy of individual users is important. To achieve data aggregation accuracy and reduction of communication overhead in wireless sensor networks is necessary in such a resource constrained network. PDA achieves an efficient data aggregation procedure called CPDA cluster based privacy data aggregation using clustering concept and algebraic properties of polynomials. A second scheme called SMART slice Mix Aggregate is also proposed based on slicing technique. The computational overhead raised in cluster based protocol is reduced in SMART at the cost of little increase in communication bandwidth. If there is no data loss in the network, both SMART and CPDA methods can result accurate aggregation in a privacy preserving way. PDA needs to be extended for specific aggregation functions.

C. People Centric Sensing System

PRISENSE [10]: To meet the privacy requirements of People centric sensing systems, the authors Shi et.al. [10] suggests a scheme based on the idea of data slicing and mixing. PRISENSE supports additive and non-additive aggregations. Three novel cover node selection strategies are used to tackle the user dynamics and dynamic nature of the network. They are random cover selection, one hop scheme and h-hop scheme.

VPA [17] is based on data slicing and mixing method. It addresses the user privacy and integrity of the data. VPA is designed for both additive and non-additive aggregation function. Here the idea is to divide the aggregation process in to two phases. In the first phase every node computes a homomorphic MAC of its original data and submits it to the aggregation server. The homomorphic property enables aggregation server to generate desired statistics without recovering the original data contributed by the user. In the second phase, using data slicing and mixing technique each user share their own data with the selected peers and then submit the mixed data to the AS. The aggregation server is now able to verify the integrity of the data shared by the user with the data submitted in the first phase. Hence VPA requires multiple rounds of bidirectional communication between the aggregation server and mobile nodes which leads to long delays. VPA is not suitable for time series data and also not fault tolerant to failure of mobile nodes. VPA+ is designed for non-additive aggregation functions through a unique combination of the binary search and verifiable privacy preserving count queries. VPA is not resistant to differential attacks.

Emiliano De Cristofaro and Roberto Di Pietro [8] focus on query and data privacy. The sensed data should be protected against unauthorized access and also the queriers might not be willing to reveal their interests. Adversarial models and strategies are discussed with the preferred dissemination method for the data. In non-resident adversary model, the adversary is not always present in the network but it corrupts after both the sensing and dissemination phase have been completed. A resident adversary is always on the network and controls the sensors at all times. Privacy is harder to achieve in the presence of resident adversary. The adversary selects the sensors to compromise based on two strategies. If the adversary is randomly distributed over the network, it controls m randomly selected sensors. Otherwise adversary focuses on a specific region of the network. Though the degree of privacy is higher in a non-resident adversary, they incur higher message overhead. The various dissemination strategies suitable for the adversarial

model are discussed and analyzed. The proposed distributed privacy preserving technique for each type of adversarial models rely on generating replica of the sensed data. Replication not only achieves privacy but also enhances data reliability and fault tolerance.

Qinghua Li et al. [14] propose an efficient protocol to achieve sum aggregate that uses the additive homomorphic encryption technique. The straw man construction algorithm for key generation is extended to reduce the computation overhead at the aggregator. The derived key management technique supports large plain text space and also achieves better security. The sum aggregate protocol is also extended to support time series data. This protocol does not require bidirectional communication between the aggregator and mobile users in every aggregation period there by reducing the communication overhead. The protocol protects the privacy of the user's data in the presence of untrusted aggregator and hence supports strong adversarial model. In mobile sensing applications, dynamic addition and removal of users may occur frequently. So, redundancy technique for assigning security parameters for users is used to address the user dynamics. The sum aggregation scheme has much less communication overhead and need to be extended to support other aggregation statistics.

D. Wireless Body Area Networks

PHDA [11]: Wireless body area network is used to monitor the user's health data in a real time manner. Since the users health data are highly sensitive and confidential , there is a need for privacy preserving data aggregation schemes to monitor the health data statistics of patients in a timely manner. PHDA uses a cloud based wireless body area network to store and process the sensed data in large scale. For privacy preserving data aggregation, bilinear map and Paillier cryptosystems are used to generate security parameters. A priority based data aggregation scheme is used to provide different forwarding strategies for the dataset with higher priorities. This not only guarantees the forwarding delay but also reduces the communication overhead. PHDA preserves both the identity and data privacy of users. It is also resistant to forgery attacks in the presence of both internal and external entity attacks.

PPM-HDA [25]: In wireless body area networks, there is a need for reliable health data aggregation in a timely manner. PPM-HDA addresses the need for a fault tolerant cloud based framework to manage the user's sensitive health data in a large scale network. Both the temporal and spatial health data statistical aggregation is taken into account. Also the scheme is implemented for additive and non-additive data aggregation schemes and offers more services in a privacy preserving way. PPM-HDA offers differential privacy and considers a strong adversary model. The additive aggregation function uses Boneh-Goh-Nissim cryptosystem which is a public key encryption scheme to protect the user's privacy. In the health care cloud, each cloud server is a powerful entity. Taking advantage of the fact, it is assumed that only a minimum set of cloud servers can be compromised by a strong adversary. The PPM-HDA scheme guarantees that the remaining uncompromised cloud servers can decrypt the aggregated data contributed by the health care sensors. Hence the scheme achieves fault tolerance and also adaptable to dynamic users. The non-additive aggregation scheme adopts prefix membership verification scheme along with binary search and count queries. The prefix membership verification scheme is used to reduce the computation overhead by changing the question of verifying whether a datum belongs to range of data into few questions of verifying whether a numerical value is equal. Also the non-additive aggregation scheme is experimented with and without achieving differential privacy. With the assistance of cloud servers the computational overhead is significantly reduced.

The design goals addressed in the related privacy preserving aggregation schemes in their corresponding domain is summarized in Table 1.

Table 1
Design goals of privacy preserving aggregation schemes

SCHEMES	*DESIGN GOALS*	*DOMAIN*
VPA[17]	Aggregation accuracy, Aggregation/data Integrity, Data/User privacy, supports additive and non-additive aggregation.	PCS
PRISENSE[10]	User privacy supports additive and non-additive aggregation.	PCS
EPPA[16]	Multidimensional data aggregation, Data confidentiality, Authentication and data integrity, Communication effectiveness.	Smart grid
PDAFT[12]	Privacy preserving, strong adversary model, Spatial and temporal aggregation, Fault tolerance	Smart grid
Qinghua Li et al.[14]	Privacy preserving, Supports sum and min for temporal data, Key management technique, Deals with dynamic joins and leaves.	Mobile sensing
PHDA[11]	Priority based aggregation, cloud assisted network, Identity and data privacy preservation.	WBAN
MUDA[13]	Privacy preserving Additive aggregation function, resist differential attacks	Smart grid
PPM-HDA[25]	Supports additive and non-additive aggregation, cloud based network, strong aggregation model, resists differential attacks	WBAN
PDA[19]	Two efficient Privacy preserving aggregation schemes, aggregation accuracy.	WSN
Emiliano De Cristofaro[8]	Query privacy and data privacy	PCS

4. CONCLUSION

With increasing growth in Internet of things and interest in social networking, people centric sensing systems are becoming popular. The privacy of the participating individuals has to be guaranteed so that the people will be willing to contribute data and that will make the system grow to a large extent. Security and users privacy are the primary factors to be addressed. This paper discusses the basic system model and threat model during the process of privacy preserving data aggregation. In wireless networks, various privacy preserving aggregation schemes has been implemented suitable to the domain. This paper focuses on the study about the design goals of the closely related privacy preserving aggregation schemes. Also the concepts behind the schemes and goals that can be addressed are identified. A PCS need to be fault tolerant to suite the user mobility and dynamic nature of the network. Any privacy preserving aggregation scheme should be adaptable to support both additive and non-additive statistical function. The adversarial model and strategies in PCS enables us to consider a strong adversary model in the design of aggregation schemes. Differential privacy is much important facts to be considered during the aggregation process. To reduce the communication and computation overhead, cloud assisted PCS systems can be vital.

References

1. A. Kapadia, D. Kotz, and N. Triandopoulos, Jan 2009, "*Opportunistic sensing: Security challenges for the new paradigm*," in *Proc. COMSNETS*.
2. A.T. Campbell, S.B. Eisenman, N. D. Lane et al., 2008, "*The rise of people-centric sensing*," *IEEE Internet Computing*, Vol. 12, No. 4, pp. 12–21.
3. A.T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A.Peterson, Aug.2006, "*People-centric urban sensing*," in *Proc. ICST WICON*.
4. A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H.Balakrishnan, S. Toledo, and J. Eriksson, 2009, "*V Track: Accurate, Energy-Aware Road Traffic Delay Estimation Using Mobile Phones*," Proc. ACM Seventh Conf. Embedded Networked Sensor Systems (SenSys '09), pp. 85-98.
5. B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, Oct. 2006, "*CarTel: A distributed mobile sensor computing system*," in *Proc. ACM SENSYS*, pp. 125–138.

6. C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, June 2008, "*Anonysense: Privacy-aware people-centric sensing*," in *Proc.ACM MobiSys*, pp. 211–224.
7. D. Westhoff, J. Girao, and M. Acharya, Oct. 2006, "*Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation*," *IEEE Trans. Mobile Comput.*, Vol. 5, No. 10, pp. 1417–1431.
8. Emiliano De Cristofaro and Roberto Di Pietro,June 2013, "*Adversaries and countermeasures in Privacy-Enhanced Urban Sensing systems*", *IEEE Systems Journal*, Vol. 7, pp. 311-322.
9. I. Krontiris, F. C. Freiling, and T. Dimitriou, 2010 "*Location privacy in urban sensing networks: research challenges and directions*," *IEEE Wireless Communications*, Vol. 17, No. 5, pp. 30–35.
10. J. Shi, Y. Zhang, Y. Liu, 2010, "*Prisense: privacy-preserving data aggregation in people-centric urban sensing systems*," *INFOCOM, 2010 Proceedings IEEE,* pp. 1-9.
11. K. Zhang, X. Liang, M. Baura, et al, 2014,"*PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs*," *Information Sciences,* Vol. 284, pp. 130-141.
12. L. Chen, R. Lu, Z. Cao, 2014, "*PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications*," *Peer-to-PeerNetworking and Applications,* pp. 1-11.
13. L. Chen, R. Lu, Z. Cao, et al, 2014, "*MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications*," *Peer-to-Peer Networking and Applications,* pp. 1-16.
14. Q. Li, G. Cao, T. La Porta, 2013, "*Efficient and privacy-aware data aggregation in mobile sensing*," *Dependable and Secure Computing, IEEE Transactions on,* Vol. 11, pp. 115-129.
15. R.K. Ganti, N. Pham, Y.-E. Tsai, and T.F. Abdelzaher, Nov. 2008, "*Poolview:Stream privacy for grassroots participatory sensing*," in *Proc. ACM SenSys*, pp. 281–294.
16. R. Lu, X. Liang, X. Li, et al, 2012, "*Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications*," *Parallel and Distributed Systems, IEEE Transactions on,* Vol. 23, pp. 1621-1631.
17. R. Zhang, J. Shi, Y. Zhang, et al, 2013, "*Verifiable Privacy-Preserving Aggregation in People-Centric Urban Sensing Systems*," *Selected Areas in Communications, IEEE Journal on,* Vol. 31, pp. 268-278.
18. S.B. Eisenman, E. Miluzzo, N.D. Lane, R.A. Peterson, G.-S. Ahn,and A.T. Campbell, 2007 ,"*The Bikenet Mobile Sensing System for Cyclist Experience Mapping*," *Proc. ACM Fifth Int'l Conf. Embedded Networked Sensor Systems (SenSys '07*), pp. 87-101.
19. W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. F. Abdelzaher, May 2007, "*PDA:Privacy-preserving data aggregation in wireless sensor networks*," in *Proc. IEEE INFOCOM*, pp. 2045–2053.
20. W. Zhang, C. Wang, and T. Feng, Mar. 2008, "*Gp2s: Generic privacy-preservationsolutions for approximate aggregation of sensor data (concise contribution)*," in *Proc. IEEE PerCom*, pp. 179–184.
21. D. Bonet, E.-J. Goh, and K. Nissim, 2005, "*Evaluating 2-DNF Formulas on Ciphertexts*," *Proc. Second Int'l Conf. Theory of Cryptography* (TCC '05).
22. L. Sweeney, 2002, "*k-anonymity: a model for protecting privacy*," *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, vol. 10, no. 5, pp. 557–570.
23. M. Welsh and J. Bers, 2010, "*CitySense: an urban-scale sensor network*," in *Ecological Urbanism*, M. Mostafavi and G. Doherty, Eds., Harvard Graduate School of Design, Lars Möuller, Zurich, Switzerland, pp. 164–165.
24. Zileng Wei, Baokang Zhao, Yujing Liu, Jinshu Su, 2013, "*PPSENSE: A novel Privacy-Preserving system in people centric sensing networks*", *IEEE International conference on communications and Networking*.
25. Song Han, Shuai Zhao, Qinghua Li, Chun-Hua Ju and Wanlei Zhou, 2015, "*PPM-HDA:Privacy-preserving and multi-functional health data aggregation with fault tolerance*", *IEEE Transactions on information forensics and security*, Issue: 99, Pages 1-1.