International Journal of Advanced Computer Science and Information Technology (IJACSIT)
Vol. 4, No. 3, 2015, Page: 74-84, ISSN: 2296-1739

www.elvedit.com

Developing CPDA schema in Privacy Preserving Data Aggregation for Wireless Sensor Networks

Authors

Sadegh Gilani
Department of Computer Engineering/ Islamic Azad University, Zahedan Branch, Iran
GilaniSadegh@Yahoo.com
Zahedan, Iran

Hadi Asharioun
Department of Computer Engineering/ Islamic Azad University, Zahedan Branch, Iran
Asharioun@sbu.ac.ir
Zahedan, Iran

Abstract

Due to high traffic load data in sensor networks, low bandwidth wireless links as well as high energy consumption for packet transfer, data aggregation techniques to acquire needed resources and energy. Data aggregation is a mechanism used in wireless sensor and VANETs networking to reduce energy consumption and extend the life of sensor nodes by sending data with stronger signals and avoid repetitive data transmission to the base station. Privacy preserving data aggregation in the network because of dynamic topologies, power limitations, memory, sensors and wireless communications media that could be eavesdropping is a major challenge in the outgoing data network contains important information and is lightweight so, security Information on these networks is very important. Privacy integration protocols aimed at preventing disclosure of confidential information to adversaries through the influence of the link or node data, so for security inevitably incur the overhead of communications and computing will be more. This Cluster-based Private Data Aggregation (CPDA) scheme could aggregate data without revealing any private information and consume fewer resources than others. Simulation results show that using the proposed algorithms, efficient data aggregation privacy of communications and computing overhead and energy consumption in wireless sensor network is improved and thus extend the life of the sensor nodes.

Key Words

wireless sensor network, Privacy preserving, Energy efficiency, Communications and Computing Overhead

I. INTRODUCTION

A wireless sensor network (WSN) is an ad-hoc network composed of small sensor nodes deployed in large numbers to sense the physical world. Wireless sensor networks have very broad application prospects including both military and civilian usage. They include surveillance [1], tracking at critical facilities [2], or monitoring animal habitats [3]. Sensor networks have the potential to radically change the way people observe and interact with their environment. Sensors are usually resource-limited and power-constrained.

They suffer from restricted computation, communication, and power resources. Sensors can provide fine-grained raw data. Alternatively, they may need to collaborate on in-network processing to reduce the amount of raw data sent, thus conserving resources such as communication bandwidth and energy. We refer to such in-network processing generically as data aggregation. In many sensor network applications, the designer is usually concerned with aggregate statistics such as SUM, AVERAGE, or MAX/MIN of data readings over a certain region or period. As a result, data aggregation in WSNs has received substantial attention.

As sensor network applications expand to include increasingly sensitive measurements of everyday life, preserving data privacy becomes an increasingly important concern. For example, a future application might measure household details such as power and water usage, computing average trends and mak making local recommendations. Without providing proper privacy protection, such applications of WSNs will not be practical, since participating parties may not allow tracking their private data. In this paper, we discuss how to carry privacy-preserving data aggregation in wireless sensor networks. In the following, we first elaborate two specific motivating applications of using wireless sensor network to carry out private data aggregation [4].

- As alluded above, wireless sensors may be placed in houses to collect statistics about water and electricity consumption within a large neighborhood. The aggregated population statistics may be useful for individual, business, and government agencies for resource planning purposes and usage advice. However, the readings of sensors could reveal daily activities of a household, such as when all family members are gone or when someone is taking a shower (different water appliances have distinct signatures of consumption that can reveal their identity). Hence we need a way to collect the aggregated sensor readings while at the same time preserve data privacy[2].
- Future in-home floor sensors, collecting weight information, are used together with shoe mounted sensors, collecting exercise-related information, in an obesity study to correlate exercise and weight loss. Aggregate statistics from those data are useful for agencies such as Department of Health and Human Services, as well as insurance companies for medical research and financial planning purposes. However, individual's health data should be kept private and not be known to other people.

From these data aggregation examples, we see why preserving the privacy of individual sensor readings while obtaining accurate aggregate statistics can be an important requirement. The protection of privacy also gives us add-on benefits including enhanced security. Consider the scenario when an adversary compromises a portion of the sensor nodes: when there is no privacy protection, the comprised nodes can overhear the data messages and decrypt them to get sensitive information. However, with privacy protection, even if data are overheard and decrypted, it is still difficult for the adversary to recover sensitive information.

Consequently, providing a reasonable guideline on building systems that perform private data aggregation is desirable. It is well-known that end-to-end data encryption is able to protect private communications between two parties (such as the data source and data sink), as long as the two parties have agreement on encryption keys. However, end-to-end encryption or link level encryption alone is not a good candidate for private data aggregation [4]. This is because:

- If end-to-end communications are encrypted, the intermediate nodes could not easily perform in-network processing to get aggregated results.
- Even when data are encrypted at the link level, the other end of the communication is still able to decrypt it and get the private data. Hence privacy is violated. Though research on privacy-preserving computation has been active in other domains including cryptography and data mining, previously-studied schemes are not readily applicable to private data aggregations in WSNs. Most of them are either not suitable for or too computational-expensive to be used in the resource-constrained sensor networks, as we will discuss in detail next section.

In summary, the problem of data privacy in WSNs can be classified in the Figure I.

From the Figure I, we know one aspect of privacy in WSNs research is data aggregation. This paper primarily focuses on privacy-preserving data aggregation. Many schemes have been developed to keep privacy of data aggregation in WSNs. such as CPDA, SMART, DADPP . In this paper, I propose a new scheme which is inspired by CPDA [3]. My scheme has lower computation overhead than CPDA and it is more secure. If there is no packet loss, this research scheme will be very effective without revealing any private information.

The rest of this paper is organized as follows. Section II summarizes the existing schemes for privacy-preserving data aggregation. Section III shows the considered system model and the requirements of privacy-preserving data aggregation are introduced .This scheme is introduced in section III. Section IV describes the simulation and results analysis. Finally, section V summarizes the paper and layout future research.

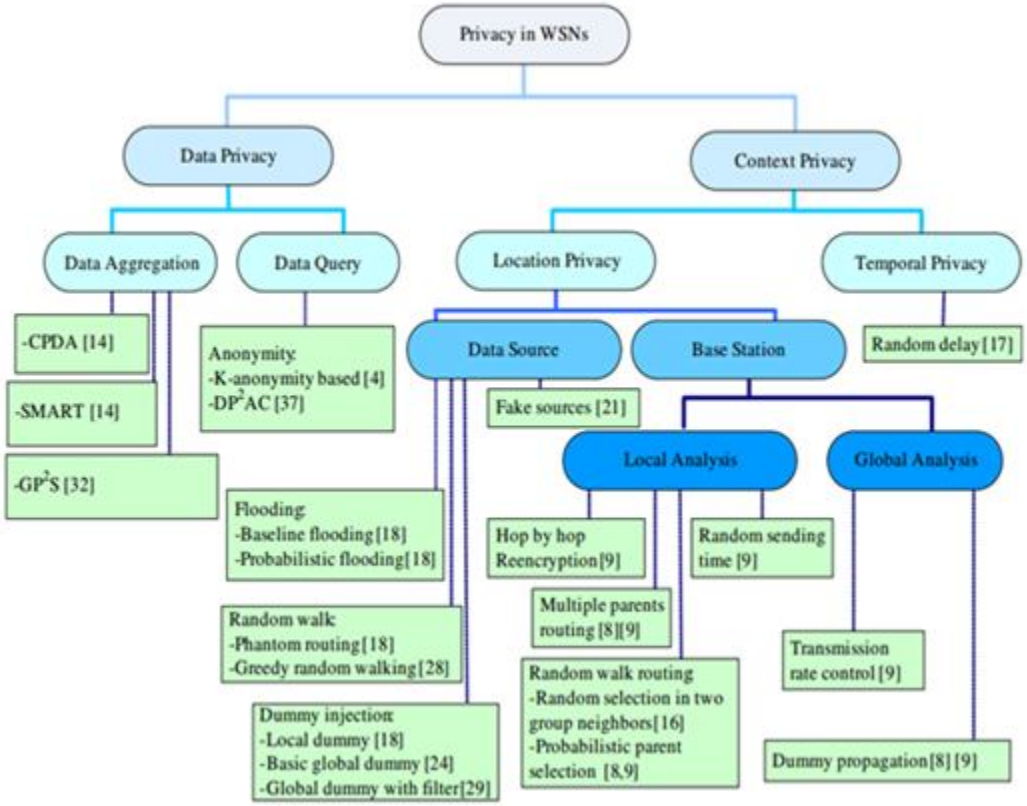

FIGURE I. TAXONOMY OF PRIVACY-PRESERVING FOR WSNS [1]

II. RELATED WORKS

In recent years, several new schemes have been proposed to solve the problem of privacy-preserving data aggregation in WSNs. Most of the existing schemes are encryption ciphers. This kind of scheme always has three steps, firstly, an intermediate aggregation node has to decrypt the received data, then aggregate the data according to the corresponding aggregation function, and finally encrypt the aggregated result before forwarding it. This scheme is fairly expensive for data aggregation in sensor networks due to sensors limited-resources.

In typical wireless sensor networks, sensor nodes are usually resource-constrained and battery-limited. In order to save resources and energy, data must be aggregated to avoid overwhelming amounts of traffic in the network. There has been extensive work on data aggregation schemes in sensor networks, including [4-9]. These efforts share the assumption that all sensors are trusted and all communications are secure. However, in reality, sensor networks are likely to be deployed in an untrusted environment, where links, for example, can be eavesdropped. An adversary may compromise cryptographic keys and manipulate the data.

Work presented in [10-12] investigates secure data aggregation schemes in the face of adversaries who try to tamper with nodes or steal the information. Work presented in [13-14] shows how to set up secret keys between sensor nodes to guarantee secure communications. For most existing secure data aggregation schemes though, an intermediate aggregation node has to decrypt the received data, then aggregate the data according to the corresponding aggregation function, and finally encrypt the aggregated result before forwarding it. This sequence is fairly expensive for data aggregation in sensor networks. To reduce computational overhead, Girao et al. [15] and Castelluccia et al. [16] propose using homomorphic encryption ciphers, which allow efficient aggregation of encrypted data without decryption involved in the intermediate nodes. Though these schemes are efficient to preserve data privacy in data aggregation, they do not protect the trend of private data of a node from being known by its neighboring nodes. This is because when the neighboring nodes can always overhear the sum of the private data and a fixed unknown number (encryption key). In contrast, the private data aggregation schemes we present in this paper ensure that no trend about private data of a sensor node is released to any other nodes.

In privacy-preservation domain, Huang, Wang and Borisov address the problem in a peer-to-peer network application in [17]. Privacy preservation has also been studied in the data mining domain [18-21]. Two major classes of schemes are used. The first class is based on data perturbation (randomization) techniques. In a data perturbation scheme, a random number drawn from a certain distribution is added to the private data. Given the distribution of the random perturbation, recovering the aggregated result is possible. At the same time, by using the randomized data to mask the private values, privacy is achieved. However, data perturbation techniques have the drawback that they do not yield accurate aggregation results. Furthermore, as shown by Kargupta et al. in [20] and by Huang et al. in [21], certain types of data perturbation might not preserve privacy well.

Another class of privacy-preserving data mining schemes [22-24] is based on Secure Multi-party Computation (SMC) techniques [25-27]. SMC deals with the problem of a joint computation of a function with multi-party private inputs. SMC usually leverages public-key cryptography. Hence SMC-based privacy-preserving data mining schemes are usually computationally expensive, which is not applicable to resource-constrained wireless sensor networks.

As we will show in the rest of this paper, unlike previous privacy-preserving approaches, our new private data aggregation schemes have the advantages: (1) They preserve data privacy such that individual sensor data is only known to their owner; (2) The aggregation result is accurate when there is no data loss; (3) They are more efficient and hence more suitable for resource-constrained wireless sensor networks.

III. PROPOSED TECHNIQUE

Due to the energy limitation problem, in-network processing becomes an important area of research in WSNs. Apart from that, the ability of providing the opportunity of complex applications running at the application layer and scalability factor makes in-network processing very attractive. Data aggregation is part of in-network processing, which is called In-Network Aggregation (INA) [13]. In most of the in-network processing use cases security and privacy issues need to be taken care of with good amount of attention [1, 7, 12]. When the requirement is like that of Yao's millionaire problem [3], where the data cannot be revealed, concept like Tinysec [14] does not work. Tinysec has the serious flaw that data has to be encrypted and decrypted at aggregator node. There are numerous practical use cases where aggregated data result is important and the individual data values are to be kept private. Consider the case of rating of television viewership, where the aggregated sum viewership result of a particular program is required by the surveying authority. But the advertisers or other third parties may be interested on the viewership details of the individual for their business interest. If these parties can access the micro details of individual viewership pattern, the privacy of the individual viewers is severely violated. In another case, an authority is responsible for billing or for resource planning an individual's water consumption in monthly basis. In the case the authorization body gets the information on the daily water consumption pattern of the households some conclusion when the house is empty (when family members are gone out) can be disclosed. This can lead to theft attempt if that data is in some malicious hands. Apart from that there are innumerable applications, where data needs to be aggregated, but the content cannot be revealed. In this paper, our attempt was to solve this kind of problem of privacy preservation when data is aggregated.

In this section, we present the system model, based on which the scheme is developed. There are N numbers of source nodes or sources which collect or produce the private data. These sources are the owners of the private data. Against the query of the service provider or the server, the sources answer the query of the server. In this process, the sources should never reveal the content of the private data, i.e., they never share the private data in raw form. They perform some data perturbation technique on the raw data, from which the server cannot understand the original content of the data. The function of the server is to aggregate the data received from n servers. It may send the aggregated data value for further processing. It is also assumed that for each source at least one source is connected. The aggregator or server node has the responsibility of data aggregation and further processing of the aggregated data. This server node has connection with N number of source nodes, which are connected with the server node through wireless links. These source nodes collect the data on its own or as per the instruction by the server node. It is assumed that the also source nodes have peer-to-peer connectivity at least with one of the nodes in order to reach the aggregator.

In this paper, I set up a model as a connected graph G (V, E), where V is the sensor nodes in the network and E is the edge between vertices. Eij Data aggregation function of this model is denotes the edge between node I and node j. The number of sensor nodes in the network is N.
f (t)=f (m (t), m (t), m (t)) , where m(t) denotes the sensor reading at time t. In this paper, I focus on additive aggregation. In [10] other aggregation functions, like average, variance and other data can be derived from summation.

In addition ,in this model I adapt the semi-honest model [11].In a semi-honest model, the sensors follows the rules of the protocol, but it is able to use what it sees in the protocol to compromise others' data privacy. This model is useful since the clients and the server can't be trusted all the time. In order to secure the data communication between sensor nodes this paper employs the key distribution. Here I follow the scheme proposed in [12] which named random key pre-distribution. This scheme consists of three phases. Firstly, there is a key pool which has M keys. Every sensor node can store N keys in itself. For each sensor node N keys are randomly selected from the key pool. This set of N keys is called the node's key ring. c p denotes the probability of two sensor nodes have at least one same key. The second phase is the key-discovery phase, in which each sensor node sends out discovery messages to find out which neighbors share a common key with itself. If two neighbors share a common key, then a secure link is setup. The discovery message is performed by Merkle puzzle policy [13], which means each sensor node issues N puzzles to its neighbors. If nodes response with correct answer, then these nodes are regarded as trusted nodes. In the last phase path-key establishment, a path-key is assigned to the pairs of sensors who do not have common keys but can be connected by multi-hop secure links.

IV. EXPERIMENTS

According to [28] I set pc=1/6 .We can see from figure 2 that when the number of nodes is large enough, the average Degree of nodes is about 12.So we set di this simulation consists of two parts. The first one is the evaluation of communication ability. I simulate the process of data aggregation and set the same parameters for this scheme and CPDA. Then I calculate the bits transmitted among the nodes and find out the average bits sent out by a sensor node. For example, if a sensor node wants to send out k which it calculates. Then the bits it sends out is b=56 +log is 12.At last I plot P and show it in figure 4. From the figure we can see most of the clusters have 3 to 7 members. Here in this paper, we simulate the conditions where a cluster has 3, 4 and 5 members. Also we make a comparison with CPDA. 2(k), where 56 bits is the header. Simulation result is following.

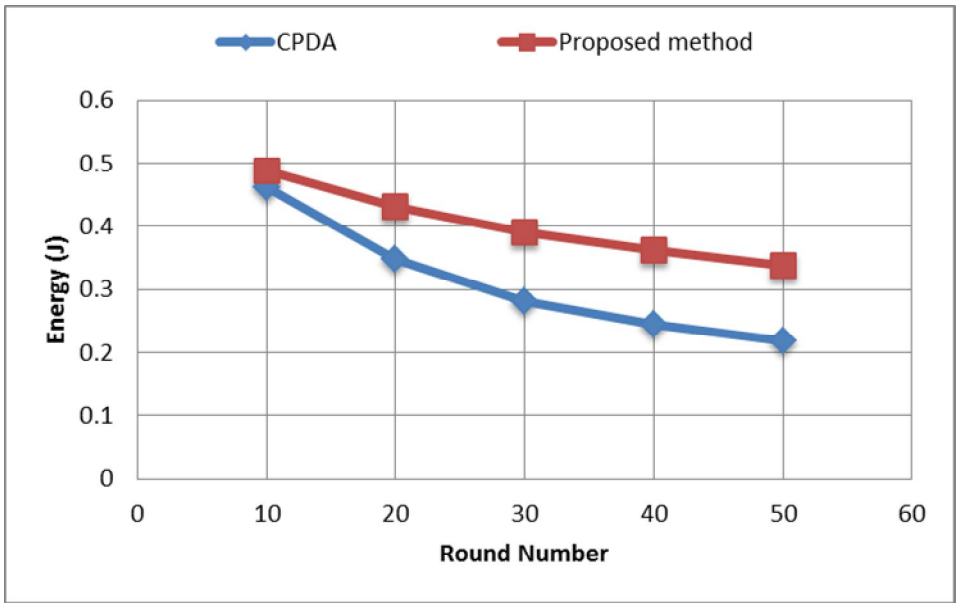

FIGURE II. VIEW NODES ENERGY CONSUMPTION

From above figure, we can see that this research scheme has similar performance with CPDA. That's because in the research scheme, complex calculation is also needed even have simplified the algorithm.

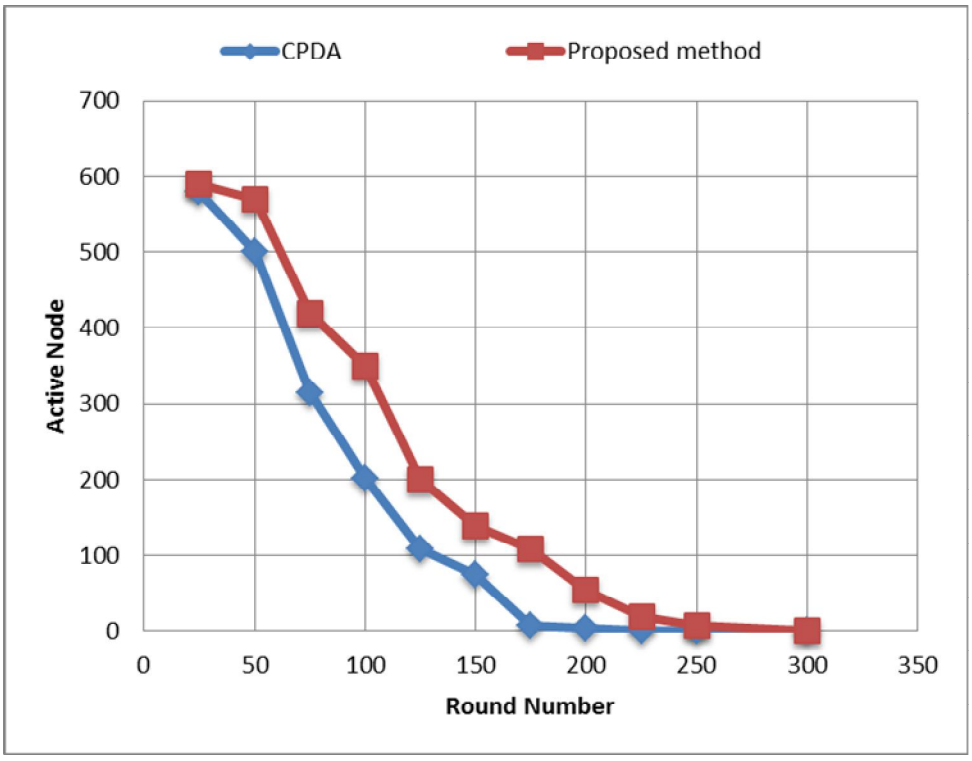

FIGURE III. LIFETIME NETWORK

From the above figure we can see that my scheme transmits less data than CPDA. In CPDA the data increase linearly with the addition of the number of the nodes in a cluster .In my scheme it increases slowly.

V. CONCLUSION & FUTURE WORKS

Providing efficient data aggregation in preserving data privacy is a challenging problem in wireless sensor networks. Many civilian applications require privacy, without which individual parties are reluctant to participate in data collection. This paper, have proposed two private-preserving data aggregation schemes – CPDA, and SMART – focusing on additive data aggregation functions. These two schemes are in terms of privacy-preservation efficacy, communication overhead, aggregation accuracy, and computational overhead.

Unlike previous privacy-preserving approaches, our new private data aggregation schemes have the advantages: (1) They preserve data privacy such that individual sensor data is only known to their owner; (2) The aggregation result is accurate when there is no data loss; (3) They are more efficient and hence more suitable for resource-constrained wireless sensor networks. Future work includes designing private-preserving data aggregation schemes for general aggregation functions. Also investigating robust private-preserving data aggregation schemes under malicious attacks.

REFERENCES

[1] Li, N. et al., 2009, Privacy preservation in wireless sensor networks: A state-of-the-art survey, Ad Hoc Networks 7.

[2] The editors of Technology Review,“10 Emerging Technologies that will change the world,” Technology Review Magazine (MIT),February,2013.

[3] D. Culler, D. Estrin, and M. Srivastava, “Overview of Sensor Networks,” IEEE Computer, Aug 2014.

[4] S. Madden, M. J. Franklin, and J. M. Hellerstein, “TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks,” OSDI, 2012.

[5] C. Itanagonwiwat, R. Govindan, and D. Estrin, “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks,” MobiCom, 2012.

[6] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, “Impact of Network Density on Data Aggregation in Wireless Sensor Networks,”In Proceedings of the 22nd International Conference on Distributed Computing Systems, 2002.

[7] Deshpande, S. Nath, P. B. Gibbons, and S. Seshan, “Cache-and-query for wide area sensor databases,” SIGMOD, 2003.

[8] Solis and K. Obraczka, "The impact of timing in data aggregation for sensor networks," ICC, 2004.

[9] X. Tang and J. Xu, "Extending network lifetime for precisionconstrained data aggregation in wireless sensor networks," INFOCOM, 2006.

[10] Przydatek, D. Song, and A. Perrig, "SIA: Secure Information Aggregation in Sensor Networks," In Proc. of ACM SenSys, 2003.

[11] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks," ACM MobiHoc, 2006.

[12] Wagner, "Resilient Aggregation in Sensor Networks," Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks, 2005.

[13] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in Proceedings of the 9th ACM Conference on Computer and Communications Security, November 2002, pp. 41–47.

[14] Liu and P. Ning, "Establishing pairwise keys in distributed sensor networks," in Proceedings of 10th ACM Conference on Computer and Communications Security (CCS03), October 2013, pp. 52–61.

[15] J. Girao, D. Westhoff, and M. Schneider, "CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks," in 40th International Conference on Communications, IEEE ICC, May 2005.

[16] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks," Mobiquitous, 2005.

[17] Q. Huang, H. J. Wang, and N. Borisov, "Privacy-preserving friends troubleshooting network," in Symposium on Network and Distributed Systems Security (NDSS), San Diego, CA, Feburary 2005.

[18] R. Agrawal and R. Srikant, "Privacy preserving data mining," in ACM SIGMOD Conf. Management of Data, 2010, pp. 439–450.

[19] Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," in Proceedings of The 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 2012.

[20] H. Kargupta, Q. W. S. Datta, and K. Sivakumar, "On The Privacy Preserving Properties of Random Data Perturbation Techniques," in the IEEE International Conference on Data Mining, November 2013.

[21] Z. Huang, W. Du, and B. Chen, "Deriving Private Information from Randomized Data," in Proceedings of the ACM SIGMOD Conference, June 2005.

[22] B. Pinkas, "Cryptographic techniques for privacy preserving data mining," SIGKDD Explorations, vol. 4, no. 2, pp. 12–19, 2002.

[23] W. Du and M. J. Atallah, "Secure multi-party computation problems and their applications: A review and open problems," in Proceedings of the 2001 Workshop on New Security Paradigms. Cloudcroft, NM: ACM Press, September 2001, pp. 13–22.

[24] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1026–1037, 2004.

[25] C. Yao, "Protocols for secure computations," in 23rd IEEE Symposium on the Foundations of Computer Science (FOCS), 1982, pp. 160–164.

[26] D. Ronald Cramer and S. Dziembowski, "On the Complexity of Verifiable Secret Sharing and Multiparty Computation," in Proceedings of the thirty-second annual ACM symposium on Theory of computing, 2000, pp. 325–334.

[27] J. Halpern and V. Teague, "Rational Secret Sharing and Multiparty Computation," in Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, 2004, pp. 623–632.

[28] Zhang, P, Gaoxi Xiao , Hwee-Pink Tan, 2013, Clustering algorithms for maximizing the lifetime of wireless sensor networks with energy-harvesting sensors, Elsevier.