International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 12 3812 – 3815

Optimized Data Aggregation Method for Time, Privacy and Effort Reduction in Wireless Sensor Network

Shanu Verma
Department of Computer Science & Engineering,
Ies College Of Technology
Bhopal, INDIA
vermashanu.verma@gmail.com

S.V Pandit
Department of Computer Science & Engineering,
Ies College Of Technology
Bhopal, INDIA
svpandit_pict@yahoo.co.in

***Abstract*—** Wireless sensor networks (WSNs) have gained wide application in recent years, such as in intelligent transportation system, medical care, disaster rescue, structure health monitoring and so on. In these applications, since WSNs are multi-hop networks, and the sink nodes of WSNs require to gather every sensor node's data, data aggregation is emerging as a critical function for WSNs. Reducing the latency of data aggregation attracts much research because many applications are event urgent. Data aggregation is ubiquitous in wireless sensor networks (WSNs). Much work investigates how to reduce the data aggregation latency. This paper considers the data aggregation method based on optimization of required time, maintain privacy while keeping lesser efforts by data aggregation in a wireless sensor network (WSN) and propose a method for the solution of the problem.

Keywords- *Wireless Sensor Network, Privacy Preserving, Data Aggregation, Power and Time Efficient.*

*****

I. INTRODUCTION

Privacy preserving data mining has become an important topic for research and applications in the last decade. Many methods have been proposed for privacy preserving data mining. We can simply divide these methods into two main categories. The fist category of methods modifies data mining algorithms so that they may carry out data mining operations on distributed datasets without knowing the exact values of the data or without directly accessing to the original datasets. Methods in the second category modify the values of the original datasets to protect their privacy.

In the second category several randomization-based data distortion methods focus on perturbing the whole dataset or the confidential parts of the dataset using certain distribution of random noises. They have been discussed in [1, 2]. Recently, matrix decomposition and factorization techniques have been used to distort numerical valued datasets in the applications of privacy-preserving data mining. In particular, SVD [3] and NMF [4] have been used for data privacy preservation and shown to be very effective in providing high level data privacy preservation and maintaining high degree data utilities. Moreover, some related signal processing techniques, for example, Fourier or wavelet transformations have also been used in data perturbation applications [5 , 6]. Both signal processing approaches seem to have a very good property of fast computation and afford good privacy protection and data utility preservation.

In a wireless sensor network (WSN), a number of sensor nodes are densely deployed in a region of interest to collect data and send the data to the data sink [7]. The data collected by different sensors usually have a certain spatial correlation and the degree of the spatial correlation increases as the distance between sensor nodes decreases [8]. The spatial correlation can lead to considerable data redundancy in the network and transmitting redundant data would cause unnecessary energy consumption. To increase energy efficiency and prolong network lifetime, it is necessary to perform in-network data aggregation to remove the data redundancy in the network.

The rest of the paper is organized as follows. Section II gives a background on the work related to this research work in privacy preservation. Section III presents the data aggregation and it's effect on the communications architecture for the WSN, as well as the different techniques for data collecting and communication through the WSN. Section IV & V describes about the literature review and proposed work respectively. Section IV describes the simulation and testbed environment used to obtain the results in this paper which are evaluated in same section. Finally, paper conclusions in Section VI.

II. PRIVACY PRESERVATION

Privacy preservation is playing an important issue while researchers discuss about any data repository in today's context of extreme penetration of network. This kind of privacy plays more vital role in the MANET where gathering of data is very frequent in network and in later part which could be very useful for processing and co-operative computations. This area researches preferred to apply various data mining techniques to preserve the privacy or secrecy of content. These techniques can also used in MANET nodes too.

There are number of techniques which are illustrated to effectively preserve the privacy of the source data. Randomization method is one of those very popular techniques. The randomization method is a technique in which noise is added to the data to be privacy-protected. This is done

to mask the attribute values of records [9, 10]. In this technique, a random noise value is added to the data which provides sufficiently large distance so that individual values cannot be recovered. After that, there are lot of techniques which are designed to get proper result from the perturbed data values. Subsequently, data mining techniques can be developed in order to work with these aggregate distributions. This randomization method has been part of traditional methods which are used in the area of perturbing data by any probability distribution method.

There are two major classes of privacy preservation schemes are applied. One is based on data perturbation techniques, in which some specific noise distribution is added to the data. Even after applying this distribution of the random perturbation, the desired result is achieved. In another technique, randomized data is used to data to mask the private values. However, data perturbation techniques have the drawback that they do not yield accurate aggregation results. It is noted by Kargupta et al. [11] that random matrices have predictable structures in the spectral domain. This predictability develops a random matrix-based spectral-filtering technique which retrieves original data from the dataset perturbing by removing or adding some random values. Perturbation is of two kinds. Additive Perturbation, in this a random noise is added to the private or personal data values. At the other end, Multiplicative Perturbation is there. In this perturbation random rotation techniques are used to get perturbed the values.

Secure Mult party Computation (SMC) and privacy preservation are related with each other with intimacy. Specially when researchers require to perform some process to hide the original data by third party. SMC problem was discussed by Yao [12] firstly, which gives a solution to Yao's Millionaire problem. SMC solution is meant for dealing with computing any function on any given input, in a distributed network in which each and every participant holds one input, which will be ensuring various things like input independence, rightness of the processing and computation, and certainly that will ensure that no information will get revealed to anyone in the computation [13].

III. DATA AGGREGATION

Due to the energy limitation problem, in-network processing becomes an important area of research in MANET 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) [14]. In most of the in-network processing use cases security and privacy issues need to be taken care of with good amount of attention [13,15,16]. When the requirement is like that of Yao's millionaire problem [12], where the data cannot be revealed, concept like Tinysec [17] 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 any of participating party can be able to find micro details of particular or individual pattern of viewership, in that case, there is severe violation of privacy of the individual viewers. 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.

IV. LITERATURE REVIEW

There are various works which have been done. Some of those work is discussed here.

A. PEPPDA (Power Efficient Privacy Preserving Data Aggregation)

Power efficient privacy preserving data aggregation method is very important in MANET because it concentrate on power preservation. In this existing protocol, it does not provide an power efficient solution for energy constrained and security required MANET. And this is just because of overhead of performing power consuming decryption and encryption at the aggregator node for the data aggregation. Ultimately, this increased number of transmissions for achieving data privacy. Aggregator node will get increased the frequency of node compromise attack because of applying decrypting algorithm at it's end. That's the reason why aggregator node is responsible to reveal large amounts of data to adversaries. The privacy homomorphism based privacy preservation protocol achieves non delayed data aggregation by performing aggregation on encrypted data. It will decrease frequency of node compromise attack. So with this way sink will get accurate aggregated results with reduced communication and computation overhead. Where time and security play an important role this technique plays very importance and vial role. Researchers have gone through all the above and come to now cited but when researchers think about its' various reason then researchers come to know that that are many reasons such as: privacy preservation, authenticity for data, data accuracy. Beside of these it also provides many other things as well like end to end confidentiality, energy efficiency and data freshness during data aggregation. This methods gives these all cited without overhead on the battery of sensors.

B. EEHA (Energy Efficient and High Accuracy secure Data Aggregation)

This [19] scheme is responsible to provides various quality aspects like high accuracy, secure data aggregation without releasing private sensor reading. This method provides all these without introducing considerable overhead on the battery limited sensors. It overcomes the issue of communication overhead by applying a slicing operation only at the leaf node.

C. Integrity-protecting Private Data Aggregation (iPDA)

This is another up gradation. iPDA [20] provides data privacy through slicing and assembling technique and at the

other end it provides integrity through redundancy by constructing a disjoint aggregation tree.

D. ESPDA (Energy-Efficient Secure Pattern Based Data Aggregation)

This technique [21] is for improving the energy efficiency with the help of sending pattern code instead of original or actual data in Wireless Sensor Networks. End to end encryption key of each node is responsible for providing privacy. It also provides confidentiality and message authentication for the data.

E. Secure Multi-party Computation (SMC)

As it name shows that it deals with the problem of a joint computation of a function. Which will take input from multi-party's private data. Usually, It is uses public-key concept of cryptography technique. And this reason make this method very expensive in computations. This cost is very unsuitable for those resources which have power-constrains like MANET or wireless sensor networks [23, 24, 25].

F. DATA PERTURBATION

This is another method of data mining to provide security to the original data. Under data perturbation method, a random number is chosen or taken from a some distribution. After choosing of that number method perform either addition/ multiplication to the data to make that converted. But still based on this perturbed data method can get the same results, while, maintaining the privacy achieved by using the randomized data to mask the private values. Another way around it has some drawbacks and that leads to inaccurate aggregation results. As shown by Kargupta et al. in [26] and by Huang et al. in [27], there are some data perturbation methods which cloud not preserve the privacy of the secrete personal data too.

V. OPTIMIZED DATA AGGREGATION METHOD FOR TIME, PRIVACY AND EFFORT REDUCTION (DAMTPER)

The proposed work is made for the wireless Sensor Network. The main purpose of the method is to find the efficient and less time consuming. The proposed algorithm is as follows:

Algorithm(data, n, m)
//Here data is generated by the network user scan //where it consist of n number of attribute and m //number of non-repeating records.

1. Data is collected at each node of network.
2. Data is aggregated on the basis of some attributes.
3. Aggregated Data is encrypted with the help of session key(key) given by the base station key (pk).
   a. key1 = mod(key, 10)
   b. c = bitshift(pk,key1)
   c. ku = mod(c, 10)
   d. encrypt = aggregated ^ ku

Fig 1: Optimized Data Aggregation Method for Time, Privacy and Effort Reduction (DAMTPER) algorithm.

VI. EXPERIMENTS AND RESULTS

Dataset: Dataset is a multi-hop wireless sensor network deployment using TelosB motes. It is taken from the The University of North Carolina at Greensboro [22]. The data consists of humidity and temperature measurements collected during 6 hour period at intervals of 5 seconds. multi-hop data is collected on 10th July 2010. Label '0' denotes normal data and label '1' denotes an introduced event. In this case steam from hot water is introduced to increase the humidity and temperature. There are total 4 data files. Each data file is from the separate mote of the WSN. Each file contains the data into following structure: i) Mote ID, ii) Humidity, iii) Temperature and iv) Label.

Researchers have analysis their proposed work with base paper [18] is done on the following parameters:

1. Data Aggregation (required to reduce size of dataset)
2. Encryption time (required to provide privacy)
3. Dissimilarity Matrix (required to find level of privacy)

After performing the base paper method the size of data got by researchers is 4690 and proposed work's size of data is 2674 number of records after aggregation, which is shown in fig 2.

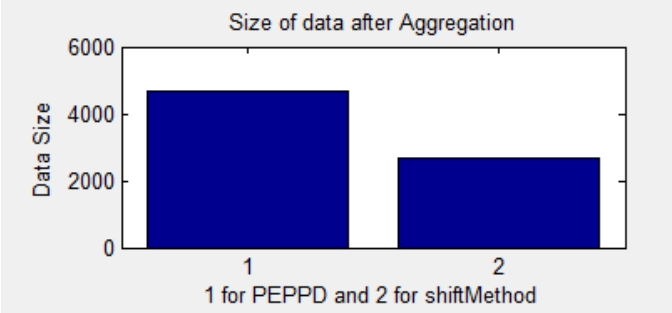

Fig 2: Comparison of Data aggregation of base [18] and proposed work.

In the similar way, when researchers' compare the base with proposed work of the basis of time require to encrypt the data or say perturb the data is 0.0911 in base paper and 0.0267 for the proposed work. Which is very clearly shown in fig 3.

In the same way, the third parameter called, Dissimilarity Matrix is is calculated for both and result is shown in fig 4.

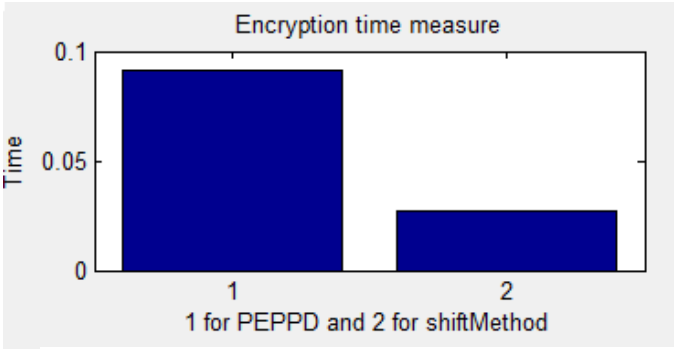

Fig 3: Comparison of required time for encryption of data through base work [18] and proposed work.

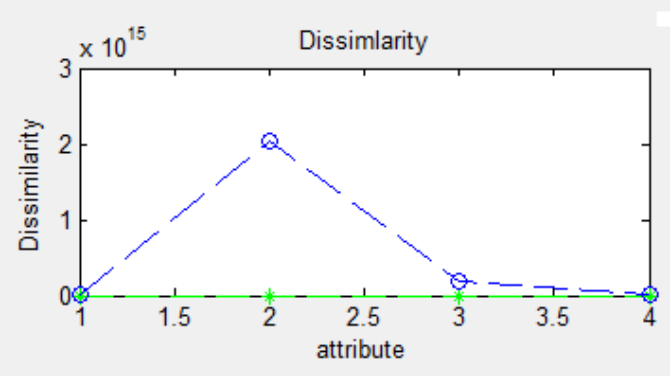

Fig 4: Comparison of dissimilarity between base work [18] and proposed work's output dataset with reference to original dataset.

VII. CONCLUSIONS

This paper investigates the Data aggregation problem with high level of efficiency in time and security, and presents an efficient algorithms Optimized Data Aggregation Method for Time, Privacy and Effort Reduction (DAMTPER). From the fig 2, 3 and 4, researchers have shown and proved that the proposed algorithm works with great efficiency over these parameters for the taken dataset.

REFERENCES

[1] Evfiievski. Randomization in privacy preserving data mining ACM SIGKDD Explorations Newsletter,4(2):43-48,2002.

[2] Z. Huang, W. Du and B. Chen. Deriving private information from randomized data. In Proceedings of the 2005 ACM SIGMOD Conference, pp. 37-48, Baltimore, MD, 2005.

[3] S. Xu,1. Zhang,D. Han and 1. Wang. Singular value decomposition based data distortion strategy for privacy protection. Knowledge and Information Systems, 10(3):383-397, 2006..

[4] J. Wang, W. 1. Zhong and 1. Zhang. NNMF-based factorization techniques for high-accuracy privacy protection on non-negative valued datasets. In Proceedings of the 2006 IEEE Conference on Data Mining, International Workshop on Privacy Aspects of Date Mining (PADM2006),pp. 513-517,Hong Kong,China,2006.

[5] S. Xu and S. Lai. Fast Fourier transform based data perturbation method for privacy protection. In Proceedings of the 2007 IEEE International Conference on Intelligence and Security Informatics, pp. 221-224,New Brunswick,NJ,2007.

[6] L. Liu,1. Wang and 1. Zhang. Wavelet based data perturbation for simultaneous privacy preserving and statistics preserving, In Proceedings of IEEE Interntional Conference on Data Mining Workshop, 2008.

[7] J. Zheng and A. Jamalipour, Wireless Sensor Networks: A Networking
Perspective, Wiley IEEE Press.

[8] T. M. Cover and J. A. Thomas, Elements of Information Theory. New
York, NY, USA: John Wiley and Sons, Inc., 1991.

[9] L. Buttyan and J. P. Hubaux, Security and Cooperation in Wireless Networks. Cambridge, U.K.: Cambridge Univ. Press, Aug. 2007.

[10] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining," ACM Sigmod, pp. 439–450, 2000.

[11] H. Kargupta, S. Dutta, Q. Wang, and K. Sivakumar, "Random-data perturbation techniques and privacy-preserving data mining," Knowledge and Information Systems, vol. 7, no. 4, pp. 387–414, 2005.

[12] A. Yao, "Protocols for secure computations," 23rd Annual Symposium on Foundations of Computer Science, pp. 160–164, 1982..

[13] S. Goldwasser, "Multi-party computations: Past and present," 16th Annual ACM symposium on Principles of distributed computing, pp. 1–6, 1997.

[14] S. Peter, K. Piotrowski, and P. Langendoerfer, "On concealed data aggregation for WSNs," IEEE Consumer Communications and Networking Conference, pp. 192–196, 2007.

[15] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks," IEEE Infocom, pp. 2045–2053, 2007.

[16] J. Deng, R. Han, and S. Mishra, "Security Support for in network Processing in Wireless Sensor Networks," ACM Workshop on Security of Adhoc Networks, pp. 83–93, 2003.

[17] C. Karlof, N. Sastry, and D. Wagner, "Tinysec: A link layer security architecture for wireless sensor networks," 2nd ACM Conference on Embedded Networked Sensor Systems, pp. 162–175, 2004.

[18] Joyce Jose, M .Princy, Josna Jose, "PEPPDA: Power Efficient Privacy Preserving Data Aggregation for Wireless Sensor Networks," IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013), pp 330-336, 2013.

[19] Hongjuan Li, Kai Lin, Kequi Li, "Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks", Computer Communication,34 (2011);591-597.

[20] W.He, X.Liu, H.Nguyen, K.Nahrstedt, T.Abdelzaher, "iPDA : An Integrity –Protecting Private Data Aggregation Scheme for Wireless Sensor Networks", IEEE MILCOM, November 2008, pp 1-7.

[21] Hassan Cam, Suat Ozdemir, Prashant Nair, Devasenapathy Muthuavinashiappan, H.Ozgur Sanli, "Energy-Efficient Secure Pattern Based Data Aggregation for Wireless Sensor Networks," Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands (ACM) 2006.

[22] http://www.uncg.edu/cmp/downloads/

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

[24] I. 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.

[25] 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.

[26] 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 2003.

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