A Survey on Secure Data Aggregation in Wireless Sensor Network

Mr.Thejaswi V
Dept of Computer Science Engg.
Akshaya Institute of Technology
Tumkur, India

Mr.Harish H K
Asst.Professor, Dept of CSE.
Akshaya Institute of Technology
Tumkur, India

***Abstract*— As wireless sensor networks grow, need of effective security mechanisms also important because sensor networks may interact with sensitive data and/or control in hostile unattended environments. In many sensor applications, the data gathered from individual nodes are combined at a home station. To cut energy consumption, many organizations also perform in-network aggregation of sensor data at intermediate nodes en route to the home station. Most of the aggregation algorithms and schemes do not include any provisions for protection, and consequently these systems are defenseless to a broad diversity of approaches. In this paper general security issues in WSNs have been explored and also present a comprehensive review of the existing literature on techniques and protocols for data aggregation in wireless sensor networks.**

Keywords: Sensor networks, aggregation, security

I. INTRODUCTION

The wireless sensor network is an ad-hoc network. It consists of small light weighted, low powered wireless nodes called sensor nodes, which are shown in Fig.1, with limited memory, computational, and communication resources [1], [2] and it measures physical parameters such as sound, force per unit area, temperature, and humidity. These sensor nodes are envisioned to play an important part in a broad diversity of fields ranging from critical military surveillance applications to forest fire monitoring and building security monitoring in the near future. In these networks, a big bit of sensor nodes are deployed to monitor a huge domain. Withal, the nodes in WSNs have severe resource constraints due to their lack of processing power, limited memory and vitality. Since these networks are commonly deployed in distant offices and left unattended, they should be fitted with security mechanisms to guard against attacks such as node capture, physical tampering, eavesdropping, denial of service, etc. Unfortunately, traditional protection mechanisms with high budget items are not feasible in resource constrained sensor nodes. The researchers in WSN security have proposed various security schemes which are optimized for these networks with resource constraints.

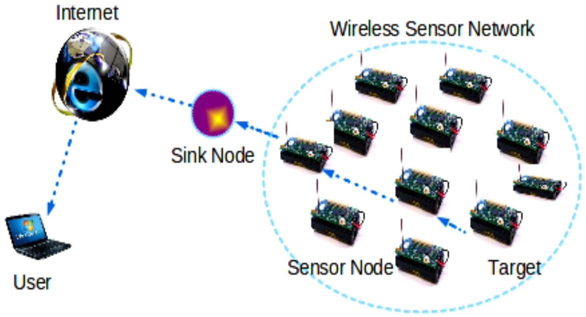

Fig.1. Sensor Network Architecture

Aim of data aggregation protocols is to combine and summarize data packets of several sensor nodes so that the amount of data transmission is reduced. An example data aggregation WSN is presented in Fig. 2 where a group of sensor nodes collects the information from a target region. When the base station queries the network, instead of sending each sensor node’s data to base station, one of the sensor nodes, called data aggregator, collects the information from its neighboring nodes, aggregates them and sends the aggregated data to the base station over a multi-hop path.

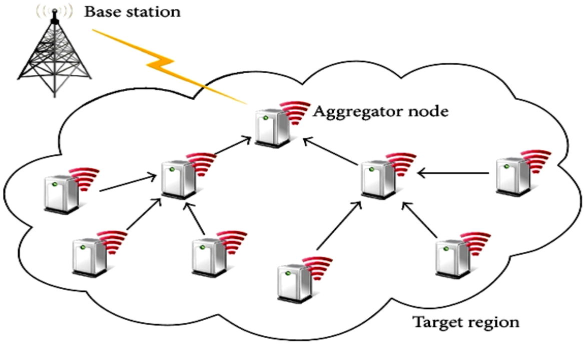

Fig.2. Data Aggregation in WSN

The rest of this paper is constructed as follows. Section II presents multicast routing problems. An overview of existing multicast-source routing algorithms is described in section III. Then, in section IV research issue is explained. Finally conclusion presented in the section V.

II. ISSUES IN DATA AGGRIGATION

A sensor network is a special type of ad hoc network. So it shares some common property of traditional networks. The security requirements of a wireless sensor network can be classified as follows:

1. Data Confidentiality: Data confidentiality is the most important issue in network security. Every network with any security focus will typically address this problem first. In sensor networks, the confidentiality relates to the following:
   - A sensor network should not leak sensor readings to its neighbors. Especially in a military application, the data stored in the sensor node may be highly sensitive.
   - In many applications nodes communicate highly sensitive data, e.g., key distribution; therefore it is extremely important to build a secure channel in a wireless sensor network.
   - Public sensor information, such as sensor identities and public keys, should also be encrypted to some extent to protect against traffic analysis attacks.

The standard approach for keeping sensitive data secret is to encrypt the data with a secret key that only intended receivers possess, thus achieving confidentiality.

2. Data Integrity and Freshness: With the implementation of confidentiality, an adversary may be unable to steal information. However, this doesn't mean the data is safe. The adversary can change the data, so as to send the sensor network into disarray. For example, a malicious node may add some fragments or manipulate the data within a packet. This new packet can then be sent to the original receiver. Data loss or damage can even occur without the presence of a malicious node due to the harsh communication environment. Thus, Data integrity guarantees that data being transferred is never been corrupted in transit.
3. Source Authentication: Even if confidentiality and data integrity are assured, we also need to ensure the freshness of each message. Informally, data freshness suggests that the data is recent, and it ensures that no old messages have been replayed. This requirement is especially important when there are shared-key strategies employed in the design. Typically shared keys need to be changed over time. However, it takes time for new shared keys to be propagated to the entire network. In this case, it is easy for the adversary to use a replay attack. Also, it is easy to disrupt the normal work of the sensor, if the sensor is unaware of the new key change time. To solve this problem a nonce, or another time-related counter, can be added into the packet to ensure data freshness.
4. Data Availability: Adjusting the traditional encryption algorithms to fit within the wireless sensor network is not free, and will introduce some extra costs. Some approaches choose to modify the code to reuse as much code as possible. Some approaches try to make use of additional communication to achieve the same goal. What's more, some approaches force strict limitations on the data access, or propose an unsuitable scheme (such as a central point scheme) in order to simplify the algorithm. But all these approaches weaken the availability of a sensor and sensor network for the following reasons:
   - Additional computation consumes additional energy. If no more energy exists, the data will no longer be available.
   - Additional communication also consumes more energy. What's more, as communication increases so too does the chance of incurring a communication conflict.
   - A single point failure will be introduced if using the central point scheme. This greatly threatens the availability of the network.

The requirement of security not only affects the operation of the network, but also is highly important in maintaining the availability of the whole network.

5. Self-Organization: A wireless sensor network is a typically an ad hoc network, which requires every sensor node be independent and flexible enough to be self-organizing and self-healing according to different situations. There is no fixed infrastructure available for the purpose of network management in a sensor network. This inherent feature brings a great challenge to wireless sensor network security as well. For example, the dynamics of the whole network inhibits the idea of pre-installation of a shared key between the base station and all sensors [3]. Several random key pre-distribution schemes have been proposed in the context of symmetric encryption techniques. In the context of applying public-key cryptography techniques in sensor networks, an efficient mechanism for public-key distribution is necessary as well. In the same way that distributed sensor networks must self-organize to support multi-hop routing, they must also self-organize to conduct key management and building trust relation among sensors. If self-organization is lacking in a sensor network, the damage resulting from an attack or even the hazardous environment may be devastating.
6. Time Synchronization: Most sensor network applications rely on some form of time synchronization. In order to conserve power, an individual sensor's radio may be turned off for periods of time. Furthermore, sensors may wish to compute the end-toend delay of a packet as it travels between two pair-wise sensors. A more collaborative sensor network may require group synchronization for tracking applications, etc. In [4], the authors propose a set of secure synchronization protocols for sender-receiver (pair-wise), multi-hop sender-receiver (for use when the pair of nodes are not within single-hop range), and group synchronization.
7. Secure Localization: The sensor network often needs location information accurately and automatically. However, an attacker can easily manipulate no secured location information by reporting false signal strengths and replaying signals, etc.

III. PRIOR STUDY WORK

Ko et al. [5] described three applications that exemplify these problems and the solutions they developed. First, they show how temporal over-sampling can simplify the analysis of a slow process such as the avian nesting cycle. Then, they show how to overcome temporal under-sampling in order to detect birds at a feeder station. Finally, they show how to exploit temporal consistency to reliably detect pollinators as they visit flowers in the field.

Corke et al. [6] concerned with the application of wireless sensor network (WSN) technology to long-duration and large-scale environmental monitoring. The Holy Grail is a system that can be deployed and operated by domain specialists not engineers, but this remains some distance into the future. They present their views as to why this field has progressed less quickly than many envisaged it would over a decade ago. They use real examples taken from their own work in this field to illustrate the technological difficulties and challenges that are entailed in meeting end-user requirements for information gathering systems. Reliability and productivity are key concerns and influence the design choices for system hardware and software.

Madden et al. [7] discussed various generic properties of aggregates, and show how those properties affect the performance of their in network approach. They include a performance study demonstrating the advantages of their approach over traditional centralized, out-of-network methods, and discuss a variety of optimizations for improving the performance and fault tolerance of the basic solution.

Zhao et al. [8] illustrated architecture for sensor network monitoring, then focus on one aspect of this architecture: continuously computing aggregates (sum, average, count) of network properties (loss rates, energy levels etc., packet counts). Their contributions are two-fold. First, they propose a novel tree construction algorithm that enables energy-efficient computation of some classes of aggregates. Second, they show through actual implementation and experiments that wireless communication artifacts in even relatively benign environments can significantly impact the computation of these aggregate properties. In some cases, without careful attention to detail, the relative error in the computed aggregates can be as much as 50%. However, by carefully discarding links with heavy packet loss and asymmetry, they can improve accuracy by an order of magnitude.

Considine et al. [9] presented new methods for approximately computing duplicate-sensitive aggregates across distributed datasets. An elegant building block which enables their techniques are the duplicate-insensitive sketches of Flajolet and Martin, which give us considerable freedom in their choices of how best to route data and where to compute partial aggregates. In particular, use of this duplicate-insensitive data structure allowed us to make use of dispersity routing methods to provide fault tolerance that would be inappropriate otherwise.

Nath et al. [10] proposed synopsis diffusion, a general framework for achieving significantly more accurate and reliable answers by combining energy-efficient multi-path routing schemes with techniques that avoid double-counting. Synopsis diffusion avoids double-counting through the use of order- and duplicate-insensitive (ODI) synopses that compactly summarize intermediate results during in-network aggregation. They provide a surprisingly simple test that makes it easy to check the correctness of an ODI synopsis. They show that the properties of ODI synopses and synopsis diffusion create implicit acknowledgments of packet delivery. They show that this property can, in turn, enable the system to adapt message routing to dynamic message loss conditions, even in the presence of asymmetric links. Finally, they illustrate, using extensive simulations, the significant robustness, accuracy, and energy-efficiency improvements of synopsis diffusion over previous approaches.

Yang et al. [11] demonstrated SDAP, a Secure Hop-by-hop Data Aggregation Protocol for sensor networks. The design of SDAP is based on the principles of divide-and-conquer and commit and attest. First, SDAP uses a novel probabilistic grouping technique to dynamically partition the nodes in a tree topology into multiple logical groups (sub trees) of similar sizes. A commitment based hop-by-hop aggregation is performed in each group to generate a group aggregate. The base station then identifies the suspicious groups based on the set of group aggregates. Finally, each group under suspect participates in an attestation process to prove the correctness of its group aggregate. Their analysis and simulations show that SDAP can achieve the level of efficiency close to an ordinary hop-by-hop aggregation protocol while providing certain assurance on the trustworthiness of the aggregation result. Moreover, SDAP is a general-purpose secure aggregation protocol applicable to multiple aggregation functions.

Yu [12] aimed to enable aggregation queries to tolerate instead of just detecting the adversary. To this end, they propose a novel tree sampling algorithm that directly uses sampling to answer aggregation queries. It leverages a novel set sampling technique to overcome a key and well-known obstacle in sampling — traditional sampling technique is only effective when the predicate count or sum is large. Set sampling can efficiently sample a set of sensors together, and determine whether any sensor in the set satisfies the predicate (but not how many). With set sampling as a building block, tree sampling can provably generate a correct answer despite adversarial interference, while without the drawbacks of traditional sampling techniques.

Roy et al. [13] showed that even if a few compromised nodes contribute false sub-aggregate values, this results in large errors in the aggregate computed at the root of the hierarchy. They present modifications to the aggregation algorithms that guard against such attacks, i.e., they present algorithms for resilient hierarchical data aggregation despite the presence of compromised nodes in the aggregation hierarchy. They evaluate the performance and costs of their approach via both analysis and simulation. Their results show that their approach is scalable and efficient.

IV. RESEARCH ISSUES

We present a comprehensive overview of secure data aggregation concept in wireless sensor networks in section I and survey on data aggregation protocols. Although the presented research addresses the many problems of data aggregation, there are still many research areas that need to be associated with the data aggregation process, especially from the security point of view.

As for the general data aggregation concept, the relation between routing mechanisms and data aggregation protocols have been well studied as they are highly correlated topics. In addition to diffusion and tree-based data aggregation protocols, many cluster-based data aggregation protocols that route aggregated data over cluster heads have been proposed. Although, these protocols shown to be very efficient in static networks in which the cluster structures do not change for a sufficiently long time, in dynamic networks they perform quite poorly. Hence, data aggregation in dynamic environments is a possible future research direction. The impact of sensor node heterogeneity over the data aggregation protocols is another unexplored research area [14]. The protocols that use powerful sensor nodes as data aggregators presented promising results. However, determining locations of these powerful nodes for the best data aggregation results needs further research.

Security is an important issue for data aggregation process and it needs to be further investigated. Clearly, there are still secure data aggregation issues that have not been addressed by the existing research. One such problem is compromised data aggregators that inject false data during data aggregation. Because data aggregation usually results in alterations in collected sensor data, false data injections by compromised data aggregators are hard to detect. There is only limited work targeting this problem and the proposed techniques are all based on extensive node monitoring mechanisms [15][16][17]. The efficiency of these node monitoring protocols is not fully evaluated and they usually incur high radio and sensing resource consumption. Hence, development of lightweight monitoring mechanisms specifically for secure data aggregation process is an interesting problem for future research.

In order to provide end-to-end security, privacy homomorphism based secure data aggregation protocols have drawn considerable attention recently. However, the design and implementation of resource efficient privacy homomorphic aggregation functions yet to be explored. Many existing public key cryptography based privacy homomorphic functions are not feasible for resource limited sensor nodes. Hence, in some secure data aggregation schemes elliptic curve cryptography is employed [18]. However, these elliptic curve cryptography based privacy homomorphic functions can only work for some specific query-based aggregation functions, e.g., sum, average, etc. Therefore, design of efficient privacy homomorphic functions that are able to work with all types of data aggregation functions needs to be explored. In addition, for certain wireless sensor network settings where real-time data delivery is demanded, symmetric key cryptography based privacy homomorphic encryption schemes are recommended. But, there are not many symmetric key based privacy homomorphic schemes. Hence, exploration of symmetric key cryptography based privacy homomorphic functions in the secure data aggregation concept is another promising research area. Using ‘‘digital watermarking” schemes to replace the expensive privacy homomorphic functions is a newly introduced concept in secure data aggregation [19]. However, this method allows only one way authentication of sensor data at the base station. Hence, investigation of two-way authentication by using watermarking techniques that will allow in-network secure data aggregation in the network may be a good research direction.

In addition, the application of source coding theory for data aggregation has drawn a little attention so far. Considering that sensor data is highly correlated, data aggregation can be achieved by employing source coding techniques. Existing research in this area focuses on only theoretical results and there are no practical algorithms applicable to wireless sensor networks yet. Moreover, there is no secure data aggregation protocol that uses the idea of source coding which may seamlessly integrate data confidentiality and aggregation together. Therefore, there is significant scope for future work in source coding based secure data aggregation.

Secure hierarchical data aggregation is expected to produce a vast amount of research in the future. Many secure data aggregation protocols assume that sensor data are aggregated at a single sink or data aggregator. Especially for privacy homomorphic secure data aggregation protocols providing hierarchical aggregation is not a trivial task. Hence, extending the current single level secure data aggregation protocols to multi layer hierarchical data aggregation protocols is an interesting problem for future research.

V. CONCLUSION

In this paper a detailed review of sensor networks and secure data aggregation concept in wireless sensor networks is provided. Behind secure data aggregation to give the motivation, first, the issues in wireless sensor networks are presented and relationship between data aggregation and security requirements are explained. Second, an extensive literature survey on data aggregation in wireless sensor network is summarized. Based on this literature survey, open research issues are given.

V. REFERENCES

[1] I.F. Akyildiz, W. Su, Y. “Sankarasubramaniam, E. Cayirci, A survey on sensor networks”, *IEEE Commun. Mag.*,Vol.8, pp. 102– 114, 2002.

[2] J. Yick, B. Mukherjee, D. Ghosal, “Wireless sensor network Survey”, *Comput. Networks*, Vol. 52, No-8, pp. 2292–2330, 2008.

[3] 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*, pp. 41–47,2002.

[4] S. Ganeriwal, S. Capkun, C.-C. Han, and M. B. Srivastava. Secure time syn- “chronization service for sensor networks”. *In WiSe ’05: Proceedings of the 4th*, 2005. ACM workshop on Wireless security, pages 97–106, New York, NY, USA, 2005.

[5] Teresa Ko, Josh Hyman, Eric Graham, Mark Hansen, Stefano Soatto, and Deborah Estrin. Embedded imagers: “Detecting, localizing, and recognizing objects and events in natural habitats”. *Proceedings of the IEEE: Special issue on sensor network applications*, Vol.98, no-11, pp. 1934–1946, 2010.

[6] Peter Corke, Tim Wark, Raja Jurdak, Wen Hu, Philip Valencia, and Darren Moore. “Environmental wireless sensor networks”. *Proceedings of the IEEE: Special issue on sensor network applications*, Vol.98, Issue-11, pp. 1903– 1917, 2010.

[7] S. Madden, M. J. Franklin, J.M. Hellerstein, and W. Hong. TAG: “A tiny aggregation service for ad hoc sensor networks”. *In Proc. of 5th USENIX Symposium on Operating Systems Design and Implementation*, 2002.

[8] J. Zhao, R. Govindan, and D. Estrin. “Computing aggregates for monitoring sensor networks”. *In Proc. of the 2nd Int’l Workshop on Sensor Network Protocols and Applications*, 2003.

[9] J. Considine, F. Li, G. Kollios, and J. Byers. “Approximate aggregation techniques for sensor databases”. *In Proc. of IEEE Int’l Conf. on Data Engineering (ICDE)*, 2004.

[10] S. Nath, P. B. Gibbons, S. Seshan, and Z. Anderson. “Synopsis diffusion for robust aggregation in sensor networks”. *In Proc. of the 2nd international conference on Embedded networked sensor systems (SenSys),* 2004.

[11] Y. Yang, X. Wang, S. Zhu, and G. Cao. SDAP: “A secure hop-by-hop data aggregation protocol for sensor networks”. *In Proc. of ACM MOBIHOC,* 2006.

[12] Haifeng Yu. “Secure and highly-available aggregation queries in largescale sensor networks via set sampling”. *In Proc. of the Int’l Conference on Information Processing in Sensor Networks,* 2009.

[13] Sankardas Roy, Sanjeev Setia, and Sushil Jajodia.” Attack-resilient hierarchical data aggregation in sensor networks”. *In Proc. of ACM Workshop on Security of Sensor and Adhoc Networks (SASN),* 2006.

[14] S. Ozdemir, Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism, in: Proceedings of the ICPS’07: IEEE International Conference on Pervasive Services, Istanbul, Turkey, pp. 165–168, 2007.

[15] B. Sun, X. Jin, K. Wu, Y. Xiao, Integration of secure in-network aggregation and system monitoring for wireless sensor networks, in: Proceedings of IEEE International Conference on Communications (IEEE ICC’07), pp. 1466–1471, 2007.

[16] B. Sun, N. Chand, K. Wu, Y. Xiao, Change-point monitoring for secure in-network aggregation in wireless sensor networks, in: Proceedings of IEEE Global Telecommunications Conference, IEEE GLOBECOM, pp. 936–940,2007.

[17] H. Çam, S. Ozdemir, False data detection and secure data aggregation in wireless sensor networks, in: Yang Xiao (Ed.), Security in Distributed Grid Mobile and Pervasive Computing, Auerbach Publications, CRC Press, 2007.

[18] D. Westhoff, J. Girao, M. Acharya, 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, 2006.

[19] W. Zhang, Y. Liu, S.K. Das, P. De, Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach, Elsevier Pervasive Mobile Comput., Vol. 4, 658– 680, 2008.