Optimized Secure Data Aggregation in Wireless Sensor Networks

Scott A. Thompson Jr. and Bharath K. Samanthula

Department of Computer Science
Montclair State University
1 Normal Ave, Montclair, New Jersey 07043
Email: {thompsons12, samanthulab}@mail.montclair.edu

Abstract— **With continuing developments in miniaturization and battery design, wireless sensor networks (WSNs) are poised to become common technology in our daily lives. Using in-network data aggregation, sensor data from multiple nodes can be combined before being forwarded to neighboring nodes; and thus, energy consumption can be reduced significantly. But in situations where sensor nodes privacy is non-negotiable, data aggregation cannot be implemented at the cost of security. Therefore, there is a strong need for secure data aggregation (SDA) protocols designed to fit the unique properties and considerable constraints of WSNs. Existing end-to-end solutions are either insecure or impractical. In this paper, we propose a novel solution for the secure aggregation of data in WSNs based on probabilistic homomorphic encryption. By combining with a unique encoding function, our solution guarantees the privacy of sensor data, while also greatly reducing communication costs.**

***Keywords*- Wireless Sensor Networks, Homomorphic Encryption, Data Aggregation**

I. Introduction

Wireless Sensor Networks (WSNs) are an attractive ambient monitoring solution for applications where traditional sensors would prove problematic [1]. Made up of large numbers of cheap, small and low-power sensor nodes, WSNs are capable of self-organizing into ad-hoc, multi-hop networks to cooperatively communicate simple sensor readings to a centralized base station. The low production cost of disposable wireless sensors and the ease of deployment make WSNs advantageous for use in environmental applications like forest fire or flood detection and military applications such as battlefield surveillance and reconnaissance [2].

Sensor nodes often have small, irreplaceable power sources. Therefore, one of the primary factors affecting network lifetime is power consumption rate [1]. A naive implementation of a WSN involves all sensor nodes directly or indirectly transmitting readings to the base station. Intermediate nodes are required to repeatedly forward readings from neighboring nodes leading to a high level of energy consumption. In general, most of the energy consumed during transmitting and receiving data is due to the power required to turn a sensor node's transceiver circuitry on and off [1]. Thus, a reduction in data transmission can yield a sizable reduction in energy consumption. An efficient method of reducing communication cost in WSNs is by implementing in-network data aggregation where the sensor nodes data is combined on the fly [3].

Confidentiality of the sensor readings are of paramount importance in many applications, such as monitoring the status and location of friendly combatants in a battle situation. In these cases, only the base station should be privy to individual sensor readings. The tack of nodes simply encrypting their readings and forwarding them to the base station may ensure security, but does so at the cost of energy efficiency. Addressing both security requirements and power constraints requires the inclusion of an efficient Secure Data Aggregation (SDA) method to the WSN [4]. Existing solutions along this direction (e.g., [5], [6]) are either insecure or less efficient.

In this paper, we propose a novel SDA solution based on the Paillier's encryption scheme [7]. Briefly, when the base station broadcasts a request to the network, each leaf node encrypts its reading and transmits it to its parent node. Included along with the sensor data, is a number indicating how many individual readings are combined in the ciphertext. The parent nodes then encrypt their own readings and aggregate them with the encrypted values received from their leaf nodes using a novel encoding technique. This is done using exponentiation and multiplication based on the homomorphic properties of Paillier [8]. Then, they sum the numbers representing the number of nodes whose data has been aggregated and forwards that along with the encrypted aggregated readings. This continues until the base station receives all of the available sensor readings in encrypted format. The base station then decrypts and decodes the data, and performs the intended function. A more detailed explanation is provided in Section III.

II. Problem Statement

Let W be a WSN with n sensor nodes, a root node, and a base station organized in a hierarchal tree [9]. Suppose

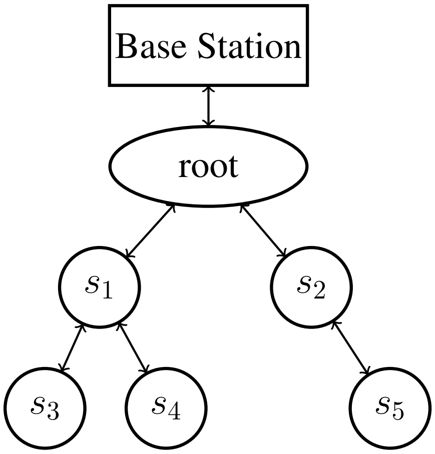

Fig. 1. A hierarchical WSN with five sensor nodes

$s_1, \ldots, s_n$ represent each nodes' corresponding sensor reading. An example of a hierarchical WSN with 5 sensor nodes is shown in Figure 1.

Under such a WSN environment, the goal is to develop an efficient SDA protocol that can deliver all of the relevant data to the base station for processing while ensuring the privacy of individual sensor nodes. That is, the base station should be the only party with access to the unencrypted sensor readings. As a result, the base station should be able to derive the results of common queries on $s_1, \ldots, s_n$ for W such as MIN/MAX, AVERAGE, COUNT, and SUM. This must be achieved while simultaneously reducing total number of packets transmitted using data aggregation.

III. Proposed SDA Protocol

In this section, we present our proposed SDA protocol using a step-by-step approach. First, we discuss our novel encoding technique that is integral to the data aggregation. Then, we discuss the general form of the protocol, along with a formal definition, and highlight how the encoding technique can be enforced under Paillier's encryption using homomorphic properties.

Let a be the encoded sensor readings and ℓ denote the uniform number of digits per reading. Without loss of generality, consider the case of encoding k sensor readings denoted by $s_1, \ldots, s_k$. To encode, the arbitrary sensor readings are placed next to each other in a systematic manner and the encoded value is given by

$$a = s'_1 \| s'_2 \| \ldots \| s'_k$$

where

$$s'_i = \begin{cases} s_i & \text{if } |s_i| = \ell \\ \underbrace{0 \ldots 0}_{l-|s_i|} \| s_i & \text{otherwise} \end{cases} \tag{1}$$

here $\|$ denotes concatenation and $|s_i|$ denotes the length of s_i in digits, for $1 \leq i \leq k$. For example, suppose $s_1 = 123$, $s_2 = 45$, and $s_3 = 789$. If $\ell = 3$, then the encoded value can be calculated as $a = \underbrace{123}_{s'_1} \| \underbrace{045}_{s'_2} \| \underbrace{789}_{s'_3} = 123045789$.

The proposed SDA protocol is comprised of three main steps described below.

Step 1: Initially, the base station broadcasts a data request to the WSN. Upon receiving the request, the leaf nodes simply encrypt their readings and send them to their parent nodes. Each sensor node transmission is a message $m_i = (\alpha_i, E(a_i))$, where α_i is the number of nodes whose readings have been included in $E(a_i)$, the encrypted aggregated sensor readings. Here $E(\cdot)$ denotes the Paillier's encryption function. Note that, for leaf nodes, $\alpha_i = 1$ and a_i denotes the corresponding sensor reading.

Step 2: Each internal node Y encrypts its own reading and proceeds as follows. Upon receiving the encrypted data from its child nodes, Y aggregates the leaf node messages. However, it is worth noting that aggregation of encrypted data is not straightforward due to the inherent overflow issue under encryption. That is, for accuracy purposes, we need to have a secure mechanism in place that can facilitate Y to check an overflow condition before aggregating the child nodes data.

In our proposed solution, we construct the overflow condition as follows. We assume that the encryption key size K is 1024 bits. Under this case, the value under encryption should be less than modulo N, where N is a part of the public key under Paillier's scheme. When $K = 1024$ bits, the value of N could be at most 2^{1024} which is roughly equivalent to $(2^{3.32})^{308.43}$ or $10^{308.43}$. Thus 308 is the largest aggregated number (in digits) that can be represented under encryption while preserving data accuracy. We emphasize that similar formulations can be deduced for other key sizes.

Definition 1. Given two messages $m_i = (\alpha_i, E(a_i))$ and $m_{i+1} = (\alpha_{i+1}, E(a_{i+1}))$, the overflow condition checks to see whether a_i and a_{i+1} can be combined inside the encryption by verifying whether their aggregated value (as per Equation 1) exceeds its maximum length. Precisely, the overflow condition for an internal node to aggregate them is $\ell * (\alpha_i + \alpha_{i+1}) < 308$, for $K = 1024$ bits. When $\ell * (\alpha_i + \alpha_{i+1}) < 308$, we say that the overflow condition holds meaning that it is safe to aggregate the two encrypted values. □

If the overflow condition holds, Y aggregates the data into a single ciphertext and adds the alpha values together. That is, we can encode both a_i and a_{i+1} under encryption by scaling and multiplying the encrypted values, $E(a)$ and $E(a_{i+1})$, together. We derived the formula for aggregating the two encrypted values, based on the previously described encoding technique (i.e., Equation 1), which is given as:

$$E((a_i * 10^{\lambda_i}) \| a_{i+1}) \leftarrow E(a_i)^{10^{\lambda_i}} * E(a_{i+1})$$

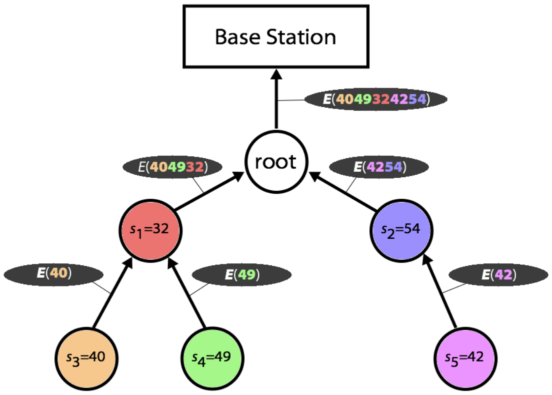

Fig. 2. Encrypted data in transmission based on the proposed SDA protocol

where $\lambda_i = \ell * \alpha_{i+1}$ represents the length (in digits) of a_{i+1}. The goal is to shift the decimal point of a_i to the right[1] the same number of digits as the length of a_{i+1}. Then we can add the adjusted a_i and a_{i+1} together and preserve the corresponding sensor readings. Due to the homomorphic property of Paillier's scheme, we can achieve the same result by exponentiation over encrypted data. By raising 10 to the power of λ_i, the decimal is moved the corresponding number of places.

The above process is repeated in a bottom-up fashion where encrypted data received at each internal node is aggregated and forwarded to their corresponding parent node in the hierarchical WSN.

Step 3: The final step of the protocol occurs once the base station has received all the encrypted aggregated values. Using the private key, it decrypts the encrypted data and decodes them into ℓ digit numbers, where each number represents the reading of a node in the WSN. Then, it performs whatever function it intended on the deduced sensor readings.

Example 1. Consider the WSN with five sensor nodes given in Figure 1. The encrypted aggregated readings that are calculated and forwarded by each node are highlighted in Figure 2. □

IV.. Performance and Security Analysis

Although we acknowledge the computational overhead of our proposed solution, it is worth noting that the increased computation cost is mitigated by the reduction in communication between internal nodes which is crucial for the long-term survival of the network. In the proposed solution, each leaf node has to perform one encryption which requires $\log N + 2$ multiplications. For an internal node with y children, the computation cost is $\alpha * \ell \log 10 + y$ multiplications, where α denotes the total number of nodes under the sub-tree rooted at the internal node.

We emphasize that the performance of the proposed solution can be improved in many ways. First, by employing widely implemented optimizations, the computation overhead on sensors can be reduced by orders of magnitude. For example, the fixed base exponentiation under Paillier's scheme can be precomputed to improve the online computation costs of each sensor. Due to space limitation, we omit the optimization details here.

On the other hand, for a 1024-bit key size, the communication cost of a leaf node is 2048 bits. Additionally, the total communication cost of an internal node with y children is $(y + 1) * 2048$ bits.

The security of the proposed scheme guarantees end-to-end data confidentiality. This is because the private key is known only to the base station and all the transmitted data is in encrypted format. As a result, the security of the proposed SDA protocol is based on the security of the Paillier's encryption scheme [7]. We emphasize that Paillier's encryption scheme provides semantic security (or IND-CPA) which is the common level of security required in most applications.

V.. Conclusion

Secure Data Aggregation (SDA) in WSNs is a growing issue in many applications, especially when the confidentiality of sensor's data is a primary requirement. In general, reducing the size of overall data communicated in a WSN is very important to the long-term survival of the network. To address this issue, we proposed a new encoding function that aggregates data within encryption in a systematic way based on the Paillier's encryption scheme. The proposed solution provides end-to-end data confidentiality and greatly reduces the communication costs. As a future work, we plan to investigate various engineering techniques to further improve the performance of the proposed solution.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci., "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] D. Puccinelli and M. Haenggi, "Wireless sensor networks: applications and challenges of ubiquitous sensing," *IEEE Circuits and Systems Magazine*, vol. 5, 2005.

[3] R. Rajagopalan and P. K. Varshney, "Data-aggregation techniques in sensor networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 8, no. 4, pp. 48–63, 2006.

[4] S. Ozdemir and Y. Xiao, "Secure data aggregation in wireless sensor networks: A comprehensive overview," *Computer Networks*, vol. 53, no. 12, pp. 2022–2037, 2009.

[5] V. Kumar and S. K. Madria, "Secure hierarchical data aggregation in wireless sensor networks: Performance evaluation and analysis," in *IEEE 13th International Conference on Mobile Data Management*, July 2012, pp. 196–201.

[6] B. K. Samanthula, W. Jiang, and S. Madria, "A probabilistic encryption based min/max computation in wireless sensor networks," in *IEEE 14th International Conference on Mobile Data Management*, 2013.

[7] P. Pallier, "Public-key cryptosystems based on composite degree residue classes," in *EUROCRYPT*, 1999, pp. 223–238.

[8] C. Fontaine and F. Galand, "A survey of homomorphic encryption for nonspecialists," *Journal on Information Security*, pp. 1–10, 2007.

[9] X. Liu, "Atypical hierarchical routing protocols for wireless sensor networks: A review," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5372–5383, Oct 2015.

[1] In general, we can move the decimal place of an integer to the right by multiplying it by powers of 10.