Data Aggregation Privacy in WSN

Jianbo Yao\textsuperscript{1} and Chaoqiong Yang\textsuperscript{2}

\textsuperscript{1}School of Mathematics and Computer, Hezhou University, Hezhou Guangxi 542899, China
\textsuperscript{2}Audit Division, Zunyi Normal College, Zunyi Guizhou 563006, China

Abstract. When WSN is applied to monitoring, the privacy of monitoring data from monitoring objects becomes an important issue for the successful application of WSN, which requires effective privacy protection for data aggregation. CPDA (Cluster-based Private Data Aggregation) has a problem that energy costs become higher with increase of nodes within a cluster. In this paper, UCPDA (Upgrade Cluster-based Private Data Aggregation) is established to perform data aggregation while preserving data privacy. According to required privacy, the sensors in a cluster are partitioned into some groups. Data preprocessing is performed only in the same group. Compared to CPDA, UCPDA has lower energy consumption for required privacy.

Keywords: WSN; Privacy; Data privacy; Data aggregation; Data aggregation privacy.

Introduction

When wireless sensor networks are applied in the field of monitoring, the privacy of monitoring data becomes a key issue for the successful application of WSN \cite{1,2}. WSN consists of a large number of sensors, which generally have limited power, computing, storage, sensing and communication capabilities. In order to save energy consumption and communication bandwidth, sensors may need to collaborate in processing fine-grained raw data collected within the network to reduce the amount of raw data sent. Data aggregation is one of the methods for in-network processing \cite{3-7}.

Without privacy protection, the data obtained by WSN can easily recover sensitive information even if the data is encrypted \cite{8,9}.

In this paper, we have created a data aggregation privacy model called UCPDA (Upgrade Cluster-based Private Data Aggregation). The model enables the sensor network to obtain accurate aggregation results while ensuring the privacy of sensor data.

The remainder of this paper is organized as follows: In Section II, direct related work. In Section III, WSN data aggregation model. In Section IV, data aggregation privacy model. In Section V, UCPDA performance analysis. In Section VI, conclusions.

Direct Related Work

Wenbo He et al. built a cluster-based private data aggregation model (CPDA), which USES the algebraic properties of polynomials to calculate the desired aggregation values and ensures that no single node knows the data values of other nodes \cite{9}. Inspired by \cite{9}, we have established a model called UCPDA (Upgrade Cluster-based Private Data Aggregation).

The CPDA model starts by building the cluster to perform intermediate aggregation. To simplify the formation of clusters, the UCPDA model establishes layered WSN. In layered WSN, each cluster head is invariant.
Second, the CPDA model performs intermediate aggregation within the cluster. For aggregation, any two nodes $A$ and $B$ exchange information within the same cluster, that is, $A$ send $\text{Enc}(v^A_b, k_{AB})$ to $B$ and $B$ send $\text{Enc}(v^B_a, k_{AB})$ to $A$. In UCPDA, nodes in the same cluster are divided into groups, and only nodes within the same group can exchange information. All preprocessed data in the same cluster is sent to the cluster header. The cluster header collects all the pre-processed data and then aggregates the data.

**Sensor Network and Data Aggregation Model**

In the UCPDA model, WSN is composed of multiple sensor nodes. Sensors are divided into three categories: base station, cluster head and common sensor nodes. The base station has abundant energy and resources; Cluster head next; sensor node is the lowest. Each node can establish a communication link with other nodes in the cluster through their different Shared key. The cluster have $n$ sensor nodes, a cluster head and $n-1$ sensor nodes. We have defined data aggregation function is as formula (1) when the sensor node ($i = 1, 2, \ldots, n$) gets sensing data $d_i$ at time $t$ [9].

$$f(t) @ f(d_1, d_2, K, d_i, K, d_n)$$

(1)

According to formula (1), the data aggregation can be calculated in layered WSN. In UCPDA, Only sum aggregate functions is focused on, because other aggregate may be reduced to sum aggregate.

**Upgrade Cluster-based Private Data Aggregation**

In order to reduce the traffic, UCPDA divides a cluster with $n$ sensor nodes into several in-network pre-processing groups according to the level of privacy protection. At the lowest level of privacy protection, the pre-processing group may include 3, 4, or 5 nodes, because $n \mod 3$ may equal 0, 1, 2. For the lowest level of privacy scenario, each group include 3 sensor nodes while $n \mod 3 = 0$. If $n = 3$, there is only a pretreatment group.

For $n = 7$, a cluster has two pretreatment groups, one is 3-member group: A, B, C held private data respective a, b, c; the other is 4-member group: E, F, G, H. held private data respective e, f, g, h. To the group with 3-member, Node A, B, C assembled values respective are

$$F_A = v^A_A + v^B_A + v^C_A = (a + b + c) + r_1x_A + r_2x_A^2$$

$$F_B = v^A_B + v^B_B + v^C_B = (a + b + c) + r_1x_B + r_2x_B^2$$

$$F_C = v^A_C + v^B_C + v^C_C = (a + b + c) + r_1x_C + r_2x_C^2$$

where $r_1 = r_1^a + r_1^b + r_1^c$, $r_2 = r_2^a + r_2^b + r_2^c$.

To the group with 3-member, first, each node uses a public non-zero number as the seed, $x_E, x_F, x_G, x_H$ which are different from each other. Then node E compute

$$v^E_E = e + r_1^E x_E + r_2^E x_E^2 + r_3^E x_E^3$$

$$v^E_F = e + r_1^E x_F + r_2^E x_F^2 + r_3^E x_F^3$$

$$v^E_G = e + r_1^E x_G + r_2^E x_G^2 + r_3^E x_G^3$$

$$v^E_H = e + r_1^E x_H + r_2^E x_H^2 + r_3^E x_H^3$$
Where \( r_1^E \), \( r_2^E \), and \( r_3^E \) are random numbers generated by node E, and known only to node E. Similar to this, node F, G, and H compute \( v_1^E \), \( v_1^F \), \( v_1^G \), \( v_1^H \); \( v_2^E \), \( v_2^F \), \( v_2^G \), \( v_2^H \) and \( v_3^E \), \( v_3^F \), \( v_3^G \), \( v_3^H \) independently as:

**Node F:**

\[
\begin{align*}
v_1^E &= f + r_1^E x_E + r_2^E x_E^2 + r_3^E x_E^3 \\
v_1^F &= f + r_1^F x_F + r_2^F x_F^2 + r_3^F x_F^3 \\
v_1^G &= f + r_1^G x_G + r_2^G x_G^2 + r_3^G x_G^3 \\
v_1^H &= f + r_1^H x_H + r_2^H x_H^2 + r_3^H x_H^3 \\
\end{align*}
\]

**Node G:**

\[
\begin{align*}
v_2^E &= g + r_1^G x_E + r_2^G x_E^2 + r_3^G x_E^3 \\
v_2^F &= g + r_1^F x_F + r_2^F x_F^2 + r_3^F x_F^3 \\
v_2^G &= g + r_1^G x_G + r_2^G x_G^2 + r_3^G x_G^3 \\
v_2^H &= g + r_1^H x_H + r_2^H x_H^2 + r_3^H x_H^3 \\
\end{align*}
\]

**Node H:**

\[
\begin{align*}
v_3^E &= h + r_1^H x_E + r_2^H x_E^2 + r_3^H x_E^3 \\
v_3^F &= h + r_1^F x_F + r_2^F x_F^2 + r_3^F x_F^3 \\
v_3^G &= h + r_1^G x_G + r_2^G x_G^2 + r_3^G x_G^3 \\
v_3^H &= h + r_1^H x_H + r_2^H x_H^2 + r_3^H x_H^3 \\
\end{align*}
\]

Then node E encrypts \( v_1^E \), \( v_2^E \), \( v_3^E \) respective using the shared key between E and F, E and G, between E and H, then sends to F, G, H, respectively.

Similarly, node F sends \( v_1^F \), \( v_2^F \), \( v_3^F \) to E, G, H; node G sends \( v_1^G \), \( v_2^G \), \( v_3^G \) to E, F, H; node F sends \( v_1^H \), \( v_2^H \), \( v_3^H \) to E, F, G.

As soon as receive \( v_1^E \), \( v_2^E \), \( v_3^E \), node E calculates assembled value

\[
F_E = v_1^E + v_2^E + v_3^E + v_1^E \\
= (e + f + g + h) + r_1^E x_E + r_2^E x_E^2 + r_3^E x_E^3
\]

where \( r_{21} = r_1^E + r_1^F + r_1^H \), \( r_{22} = r_2^E + r_2^F + r_2^G + r_2^H \), \( r_{23} = r_3^E + r_3^F + r_3^G + r_3^H \).

Similarly, node F, G, and H calculates assembled value

\[
F_F = v_1^F + v_2^F + v_3^F + v_1^F \\
= (e + f + g + h) + r_1^F x_F + r_2^F x_F^2 + r_3^F x_F^3
\]
Next, node A, B, C, E, F, G, H respective sends $F_A, F_B, F_C, F_E, F_F, F_G, F_H$ to cluster head. After cluster head knows all the assembled values: $F_A, F_B, F_C, F_E, F_F, F_G, F_H$, cluster head gets the aggregate value.

On the basis of all the assembled values: $F_A, F_B, F_C, F_E, F_F, F_G, F_H$, Formula (2) is constructed as follow:

$$
F_G = v_G^E + v_G^F + v_G^G + v_G^H
= (e + f + g + h) + r_2 x_G^1 + r_2 x_G^2 + r_2 x_G^3
$$

$$
F_H = v_H^E + v_H^F + v_H^G + v_H^H
= (e + f + g + h) + r_2 x_H^1 + r_2 x_H^2 + r_2 x_H^3
$$

respectively.

Performance Analysis
In this section, the UCPDA and the CPDA In this section, UCPDA and CPDA are compared in terms of privacy, efficiency, and accuracy.

1.1. Privacy Protecting
In UCPDA, a node can build links with the others by their shared keys, probability of privacy violation by overhear is zero. Consequently, the collusion attacks is considered only.

To prevent collusion, a group has the least 3-member, the privacy levels may be defined based on size of groups. The least privacy level is a 3-member group, secondly is a 4-member group, the rest may be deduced by analogy, the most privacy level is that the member of a group equal to the member of a cluster.

From this defined, grouping is significance only when the member of a cluster is more than 4-member. The privacy protecting levels are determined by the least grouping of a cluster.
Figure 1. Privacy Protecting Performance.

Figure 1 compare UCPDA with CPDA for privacy protecting performance while the cluster has twenty nodes. UCPDA can provide for different privacy protecting while required privacy levels. CPDA only provide for the most privacy protecting level and a larger communication overhead.

1.2. Efficiency

Figure 2 compare UCPDA with CPDA about efficiency when a cluster has 20-member.

In UCPDA, a node only communicates with the others except sending pretreated data to cluster head, but in CPDA, a node communicates with each other in a cluster, the CPDA’ overhead is larger than UCPDA’.

In Figure 2, owing to unbalanced grouping, the UCPDA’ unconventionality for the efficiency has appeared at privacy levels 5; 6; 7; 8. At privacy level 5, twenty nodes are partitioned two groups, one has 7-member and the other has 13-member. Similarly, at privacy 6; 7; 8, twenty nodes are partitioned two groups, one has 8-member, 9-member, 10-member and the other has 20-member, 11-member, 10-member, respectively. The ratio of the overhead in the 13-member group and in the 12-member group is larger than the ratio of the overhead in the 8-member group and in the 7-member group. The rest cases are similar with the case.

1.3. Accuracy

Both UCPDA and CPDA yield accurate aggregate results.

Conclusions

WSN is applied in the field of monitoring, and the data privacy of monitored objects has become an important issue. This paper has a data aggregation privacy protection model - UCPDA. In UCPDA, all nodes are partitioned into some groups while required privacy. Compared to CPDA, UCPDA has lower energy costs when required privacy.

Acknowledgements

This research was supported by the Doctor Fund project of Hezhou University under grant HZUBS201809.
References

[1] Yi, Xun, Bouguettaya, Athman, Georgakopoulos, Dimitrios, Song, Andy, Willemsen, Jan. Privacy Protection for Wireless Medical Sensor Data [J]. IEEE Transactions on Dependable and Secure Computing, 2015: 1-1.

[2] Singh, Vishal Krishna, Verma, Saurabh, Kumar, Manish. Privacy Preserving In-network Aggregation in Wireless Sensor Networks[J]. Procedia Computer Science, 2016, 94: 216-223.

[3] de Fuentes J. M., González-Manzano L., Mirezi O. Privacy Models in Wireless Sensor Networks: A Survey [J]. Journal of Sensors, 2016, 2016: 1-18.

[4] Guo, Longjiang, Li, Yingshu, Cai, Zhipeng. Minimum-latency aggregation scheduling in wireless sensor network [J]. Journal of Combinatorial Optimization, 2016, 31(1): 279-310.

[5] Ma, Lili, Liu, Jiangping, Luo, Jidong. Method of Wireless Sensor Network Data Fusion [J]. International Journal of Online Engineering (iJOE), 2017, 13(09)

[6] Anna Devi E., Martin Leo Manickam J. Identifying Partitions in Wireless Sensor Network[J]. International Journal of Parallel Programming, 2018

[7] Bhasin, Vandana, Kumar, Sushil, Saxena, P. C., Katti, C. P. Security architectures in wireless sensor network [J]. International Journal of Information Technology, 2018

[8] Raja Manjula, Datta Raja. Efficient aggregation technique for data privacy in wireless sensor networks [J]. IET Networks, 2018, 7(5): 287-293.

[9] He W, Liu X, Nguyen H, Nahrstedt K, and Abdelzaher T. "PDA: Privacy-preserving Data Aggregation in WSN", 26th Annual IEEE Conference on Computer Communications IEEE INFOCOM 2007, Anchorage, Alaska, May 2007.