Homomorphic Hash Based Scheme for Relibility Enhancement in Data Aggregation

Feng Xie*
Dept. of Electrical Engineering
Naval University of Engineering
Wuhan, China, 430074
E-mail: xiefengnue@163.com

Xiaohui Ye
Dept. of Electrical Engineering
Naval University of Engineering
Wuhan, China, 430074
E-mail: whyxh007@yahoo.com.cn

Abstract—Currently Internet of Things has been envisioned as a key technology for smart city, intelligent transportation, and green buildings. Data aggregation is largely applied to sense the data collection in Internet of Things, as it can improve the efficiency of data sensing and forwarding to final sink nodes. The reliability and efficiency of data aggregation have significant influence on the functionality of environmental parameter sensing and emergence responses. However, data reliability is usually verified by each forwarding hop for confirming the data integrity of receiving data. This verification method at each intermediate aggregative node induces much computation overhead. In this paper, researchers propose a novel scheme that requires verification only once and only at final sink node. It is a homomorphic hash function based data reliability verification scheme that largely reduces the computation overhead by only computing hash function only once. This scheme relies on the dedicated characteristic of homomorphic hash function. The theoretic analysis and simulation results justified that the scheme can improve the efficiency of data aggregation in Internet of Things.

Keywords-Homomorphic Hash; Internte of Things, Data Aggregation; Relibility; Efficiency

I. INTRODUCTION

Currently, Internet of Things (IoT) technology has been envisioned as a promising approach and has attracted more and more attentions. IoT can collect related data about surrounding environments and aggregate those data into a central server for final decision, which can help monitor the environmental parameters and fast respond to some emergent situations instantly. For example, IoT can be widely applied to smart city, intelligent transport system, and green buildings.

Data aggregation is a critical issue for sensing data collection in IoT. Sensed data will be aggregated at aggregated nodes on forwarding paths, so that the communication overhead will be largely reduced due to the length shortening of sensing data. The aggregated function could be summation and average, which can reveal the summation of the current monitored number, or average temperature. Those aggregated results are calculated at forwarding nodes instead of final sink nodes, thus the communication overhead of sensing data forwarding is largely mitigated.

In data aggregation, the reliability of aggregated data is of surmount importance. In other words, the collected data should not be modified by adversaries. The receiver can guarantee the soundness of the aggregated data. However, the reliability of data aggregation imposes some challenges due to the complexity of data aggregation. For example, the data aggregation topology is dynamically generated, thus the verification of reliability should be flexible. The data aggregation is usually conducted at forwarding nodes, thus the computation overhead for the verification should be lightweight.

In this paper, researchers propose a Homomorphic hash function based data aggregation verification method, which can be flexible and lightweight. The contribution of this paper has two folders: (1)Researchers Researchers give a general model for improving data aggregation efficiency in Internet of Things. (2)Researchers Researchers propose a homomorphic hash function based scheme for reliability check in data aggregation.

The reminder of the paper is organized as follows: the Problem Formulation section will address the normalized problem on efficient data aggregation of sensing data. The state-of-the-art is given in Related Work section. The Proposed Scheme section proposes dedicated homomorphic hash function based scheme. The Analysis section will give the extensive analysis. The Conclusion section concludes the paper.

II. PROBLEM FORMATION

Researchers here concentrate on the randomly failure in the communication channel during data aggregation.

Suppose sensing data is SD1, SD2, …, SDn. The aggregation node will receive those data and computes aggregated data, denote as AD. AD is the function of SD1, …, SDn. After the aggregation of SDi (i=1,…,n), the aggregation of AD could be possible.Researchers Researchers call it multiple aggregations.

The adversary model in this paper is data modification, data loss, and forged data injection.

The reliability of once aggregation means that after the aggregated nodes receive the sensing data, the node can guarantee the sensing data is not hacked by adversaries.

The reliability of n-time aggregation means that after the data have been aggregated for n times at n aggregation nodes, the received node can guarantee the aggregated data is not hacked by adversaries.

Researchers thus call the data aggregation reliability problem as DARP, in which the reliability of n-time aggregation data should be guaranteed.

III. RELATED WORK

Data aggregation is a key problem in Internet of Things or wireless sensor networks. M. Bagaa et. al [1] reviewed current literature in this topic. J. Lin et. al [2] proposed a evolutionary game –based data aggregation model that could tradeoff different influence parameters. Xiaohua Xu et. al [3] proposed a delay-efficient algorithm in multihop wireless sensor networks, which had similar scenario with our discussion.

Some papers explore the data aggregation algorithms [4-11] for different applications, which apply intelligent mechanism such as soft computing and machine learning. S. Srinivasan et. al [12] discussed thoroughly about survivable data aggregation, which is related to data reliability. H. Salarian et. al [13] proposed to utilize mobile sink to improve the energy efficiency. Their methods have some application limitations. D. Takaishi et. al [14] discussed big data gathering in distributed sensor networks, in which the data volume was large and reliability was not a concentration. Wenbo Zhao et. al [15] proposed to luse dynamic traffic patterns to help schedule data collection procedure so as to improve the energy efficiency. Those schemes do not concentrate on data reliability problem and do not rely on cryptographic methods.

IV. PROPOSED SCHEME

A. Research Challenges and Solution Rationale

Before the discussion of the proposed scheme, researchers pinpoint some challenges in the design.

The once aggregation problem is simple, thus it should be solved firstly. The straightforward method for data reliability is integrity checks or hash functions, here researchers concentrate the efficiency of the solution.

The difficulty of n-time aggregation is that how to guarantee the reliability of each aggregation. Although it can be done by every aggregation nodes, it has to assume the trustworthiness of each intermediate aggregation nodes. Moreover, it has to conduct the verification computation at each intermediate aggregation node, which induces more computation overhead. Thus, the problem can be narrowed to a new direction - finding a more efficient method, in which the reliability can be guaranteed by only once check.

B. Basice Scheme

Before giving the final proposed scheme, researchers discuss the basic scheme for better understanding the motivation.

Regarding the verification of once aggregation, the method can be given in the following.

The sensing nodes send following data to aggregation node:

$$SDi,\ Hash(SDi \parallel K),\ i=1,\ldots,n \qquad (1)$$

where, k is a shared secret key.

The aggregation node will check the hash function and verify the integrity of SDi.

This is straightforward method, but in case SDi is not a long bit string, the method is not efficient.

The hash function can be replaced by other efficient coding function, such as CRC function, parity function, etc.

The packet format can be denoted as

$$SDi,\ Code(SDi \parallel K),\ i=1,\ldots,n \qquad (2)$$

where, code() is a coding method that can check the integrity of SDi.

Regarding the situation of multiple aggregation, which is very likely in IoT, the verification can be done at each aggregation node. The repeat of previous procedure can check the integrity at each aggregation procedure.

C. Advanced Scheme

Researchers propose a method relying homomorphic hash function. The homomorphic hash function has following advantages:

(1) Hash(X o Y) = Hash(X) o Hash(Y), where o denotes an operation. It could be addition operation, which is determined by instantiated aggregated function. The addition is the most frequently used operation. The same operation is usually used for all n times of aggregation.

(2) The aggregation of the hash function can be used for the verification of the aggregated data. In other words, the hash function can be aggregated together with data aggregation.

Hereby, researchers give the details of the proposed scheme as follows:

(1) After sensing data, the sensing node sends:

$$SDi,\ Hash(SDi \parallel K),\ i=1,\ldots,n \qquad (3)$$

where, k is a shared secret key, and Hash() is homomorphic hash function.

(2) After receiving the packets, intermediate aggregation nodes compute:

$$AD = Agg_1(SD1,\ SD2,\ \ldots,\ SDn), \qquad (4)$$

$$Hash(AD) = Agg_1(Hash(SDi \parallel K)),\ i=1,\ldots,n \qquad (5)$$

where, $Agg_1()$ is an aggregation function. After the data aggregation, checks the soundness of Hash(AD).

(3) Multiple aggregation of multiple aggregated data.

$$AD = Agg_2(AD1,\ AD2,\ \ldots,\ ADm), \qquad (6)$$

$$Hash(AD) = Agg_2(Hash(ADi \parallel K)),\ i=1,\ldots,m \qquad (7)$$

where, $Agg_2()$ is the second aggregated function.

(6) Repeat (5) for n times,

$$AD = Agg_n(AD1,\ AD2,\ \ldots,\ ADm), \qquad (8)$$

$$Hash(AD) = Agg_n(Hash(ADi \mid\mid K)),\ i=1,\ldots,m \qquad (9)$$

where, $Agg_n()$ is the n-time aggregated function.

After n times aggregation, the final received node checks where the hash value is satisfied.

V. SIMULATION AND ANALYIS

Before the simulation verification, researchers conduct theoretic analysis.

Statement 1 The efficiency of advanced scheme can be guaranteed by once verify for all aggregation results.

Theorem 1 The computation overhead in advanced scheme is better than basic scheme in a ratio $(1/m)^n$.

Proof. Suppose the number of data to be aggregative is m, thus aggregated node will compute hash function for m times. However, if homomorphic hash function is applied, only one time of hash computation is required. The aggregation computation for hash function is usually lower than hash function in terms of computation overhead. Thus, the energy cost is 1/m of basic scheme. For n times of aggregation, the energy cost is about $((1/m)^n$ of basic scheme.

Researchers conduct the numeric simulation. The results are shown in Fig. 1-4. In those figures, energy efficiency becomes higher with the increasing of aggregative times and the increasing of aggregative nodes. Normally, researchers suggest the number of aggregative nodes is 3-5, and the layer of aggregation level is 3-5.

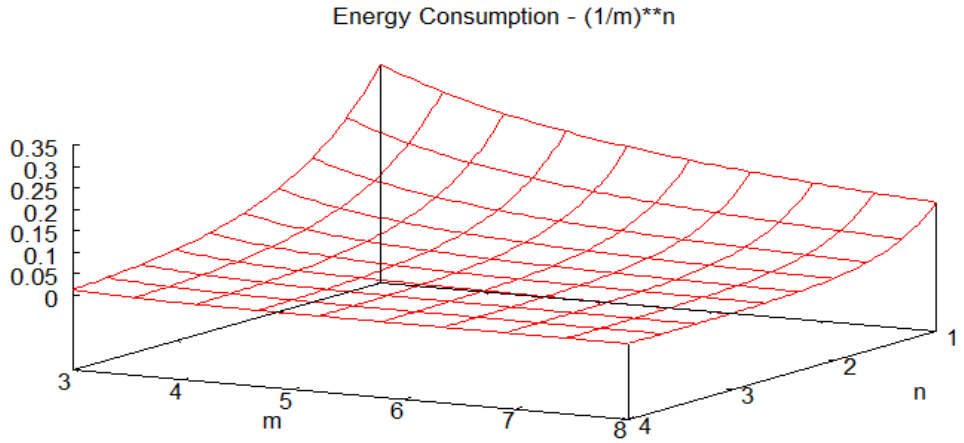

Figure 1. Numeric Simulation Result for m=[3,8],n=[1,4]

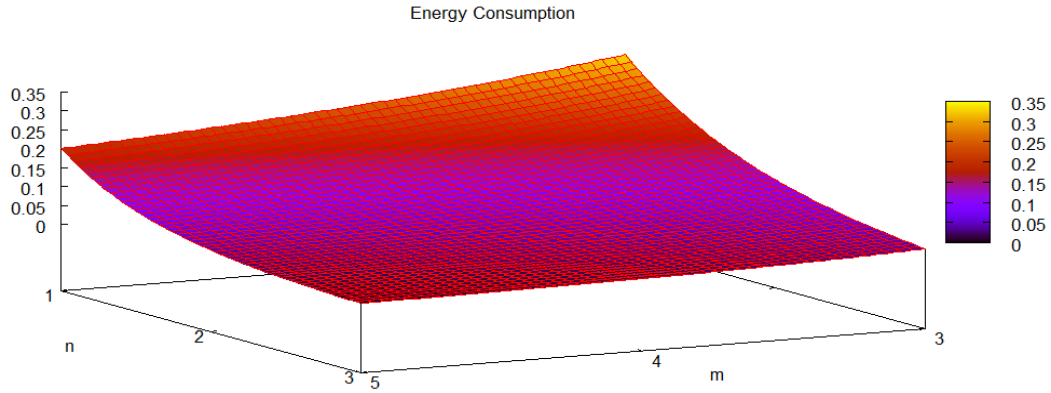

Figure 2. Numeric Simulation Result for m=[3,5],n=[1,3]

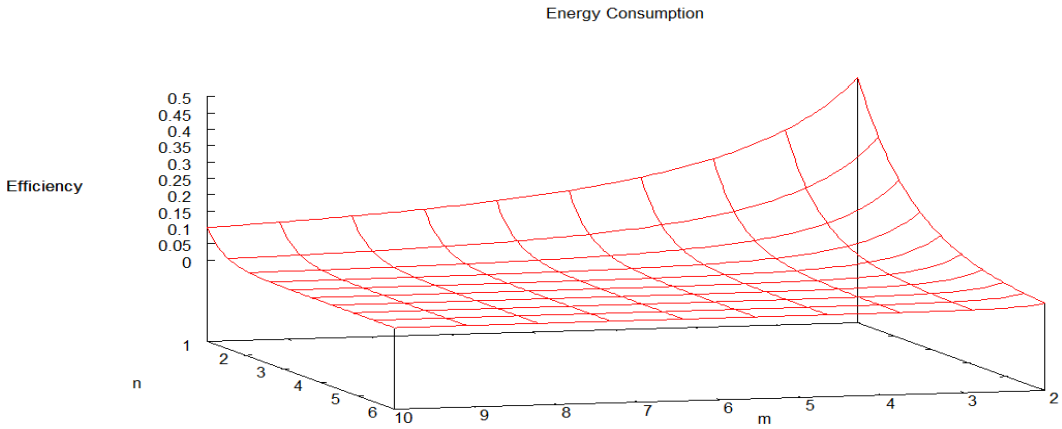

Figure 3. Numeric Simulation Result for m=[2,10],n=[1,6]

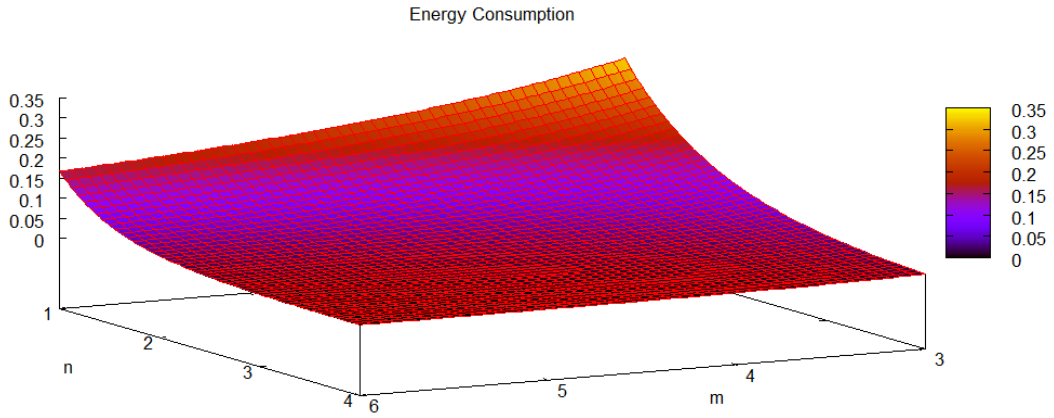

Figure 4. Numeric Simulation Result for m=[3,6],n=[1,4]

VI. CONCLUSIONS

After giving the basic model for data reliability of data aggregation, researchers firstly figure out two basic schemes that are straightforward and have low efficiency in terms of computation. Researchers then propose an advanced scheme by using homomorphic hash function that can avoid the verification computation at intermediate nodes and require to compute hash function only once at final sink node. The analysis justified that our scheme has higher efficiency. The energy efficiency becomes higher with the increasing of aggregative times and the increasing of aggregative nodes. Normally, researchers suggest the number of aggregative nodes is from 3 to 5, and the layer of aggregation level is between 3 and 5.

REFERENCES

[1] Bagaa, M.; Challal, Y.; Ksentini, A.; Derhab, A.; Badache, N., Data Aggregation Scheduling Algorithms in Wireless Sensor Networks: Solutions and Challenges, IEEE Communications Surveys & Tutorials, vol.16, no.3, pp.1339-1368, 2014

[2] Lin, J.; Xiong, N.; Vasilakos, A.V.; Chen, G.; Guo, W., Evolutionary game-based data aggregation model for wireless sensor networks, IET Communications, vol.5, no.12, pp.1691-1697, August 12 2011

[3] XiaoHua Xu; Mo Li; XuFei Mao; Shaojie Tang; ShiGuang Wang, A Delay-Efficient Algorithm for Data Aggregation in Multihop Wireless Sensor Networks, IEEE Transactions on Parallel and Distributed Systems, vol.22, no.1, pp.163-175, Jan. 2011

[4] Hoang, D.C.; Kumar, R.; Panda, S.K., Optimal data aggregation tree in wireless sensor networks based on intelligent water drops algorithm, IET Wireless Sensor Systems, vol.2, no.3, pp.282-292, September 2012

[5] Hongbo Jiang; Shudong Jin; Chonggang Wang, Prediction or Not? An Energy-Efficient Framework for Clustering-Based Data Collection in Wireless Sensor Networks, IEEE Transactions on Parallel and Distributed Systems, vol.22, no.6, pp.1064-1071, June 2011

[6] Jing He; Shouling Ji; Yi Pan; Yingshu Li, Constructing Load-Balanced Data Aggregation Trees in Probabilistic Wireless Sensor Networks, IEEE Transactions on Parallel and Distributed Systems, vol.25, no.7, pp.1681-1690, July 2014

[7] Tan, H.O.; Korpeoglu, I.; Stojmenovic, I., Computing Localized Power-Efficient Data Aggregation Trees for Sensor Networks, IEEE Transactions on Parallel and Distributed Systems, vol.22, no.3, pp.489-500, March 2011

[8] Yajie Ma; Yike Guo; Xiangchuan Tian; Ghanem, M., Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks, IEEE Sensors Journal, vol.11, no.3, pp.641-648, March 2011

[9] Mo Li; Yajun Wang; Yu Wang, Complexity of Data Collection, Aggregation, and Selection for Wireless Sensor Networks, IEEE Transactions on Computers, vol.60, no.3, pp.386-399, March 2011

[10] Ebrahimi, D.; Assi, C., A Distributed Method for Compressive Data Gathering in Wireless Sensor Networks, IEEE Communications Letters, vol.18, no.4, pp.624-627, April 2014

[11] Abu Alsheikh, M.; Shaowei Lin; Niyato, D.; Hwee-Pink Tan, Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications, IEEE Communications Surveys & Tutorials, vol.16, no.4, pp.1996-2018, Fourthquarter 2014

[12] Srinivasan, S.; Azadmanesh, A., Survivable Data Aggregation in Multiagent Network Systems with Hybrid Faults, IEEE Transactions Computers, vol.62, no.10, pp.2054-2068, Oct. 2013

[13] Salarian, H.; Kwan-Wu Chin; Naghdy, F., An Energy-Efficient Mobile-Sink Path Selection Strategy for Wireless Sensor Networks, IEEE Transactions on Vehicular Technology, vol.63, no.5, pp.2407-2419, Jun 2014

[14] Takaishi, D.; Nishiyama, H.; Kato, N.; Miura, R., Toward Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks, IEEE Transactions on Emerging Topics in Computing, vol.2, no.3, pp.388-397, Sept. 2014

[15] Wenbo Zhao; Xueyan Tang, Scheduling Sensor Data Collection with Dynamic Traffic Patterns, IEEE Transactions on Parallel and Distributed Systems, vol.24, no.4, pp.789-802, April 2013 .