IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

Data Aggregation Mechanism For Distributed In Wireless Sensor Network

Dr. D. Francis Xavier Christopher M.Sc, Mphil (CS), PGDPM&IR, Ph.D[1]., C.Akalya, MCA[2],

Director, School Of Computer Studies, Rathnavel Subramaniam College Of Arts & Science, Sulur, Coimbatore[1]

M.Phil (CS) Research Scholar, Rathnavel Subramaniam College Of Arts & Science, Sulur, Coimbatore[2]

Abstract: This paper proposes an efficient protocol to obtain the Sum aggregate, which employs an additive homomorphic encryption and a novel key management technique to support large plaintext space. The paper also extends the sum aggregation protocol to obtain the Min aggregate of time-series data. It shows that the proposed protocols are faster than existing solutions, and it has much lower communication overhead. In addition, the paper proposes a new concealed data aggregation scheme which is homomorphic public encryption system based. The proposed scheme has three contributions. First, it is designed for a multi-application environment. The base station extracts application-specific data from aggregated ciphertexts. Next, it mitigates the impact of compromising attacks in single application environments. Finally, it degrades the damage from unauthorized aggregations. In addition, Database as a Service model is proposed in which, a client stores her database on an untrusted service provider. Therefore, the client has to secure their database through Privacy Homomorphism (PH) schemes because PH schemes keep utilizable properties than standard ciphers. Based on PH schemes, the provider can conduct aggregation queries and retrieve results.

Index Terms: Data Aggregation, Centralized Approach, In-Network Aggregation, Tree Based Approach, Cluster-Based Approach

I. INTRODUCTION

Aggregation statistics need to be periodically computed from a stream of data contributed by mobile users [1], to identify some phenomena or track some important patterns in many scenarios. For example, the average amount of daily exercise (which can be measured by motion sensors [2]) that people do can be used to infer public health conditions. The average or maximum level of air pollution and pollen concentration in an area may be useful for people to plan their outdoor activities. Other statistics of interests include the lowest gasoline price in a city, the highest moving speed of road traffic during rush hour, and so on.

Although aggregation statistics computed from time-series data are very useful, in many scenarios the data from users are privacy-sensitive, and users do not trust any single third-party aggregator to see their data values. For instance, to monitor the propagation of a new flu, the aggregator will count the number of users infected by this flu. However, a user may not want to directly provide her true status ("1" if being infected and "0" otherwise) if she is not sure whether the information will be abused by the aggregator. Accordingly, systems that collect users' true data values and compute aggregate statistics over them may not meet users' privacy requirement [1]. Thus, an important challenge is how to protect the users' privacy in mobile sensing, especially when the aggregator is untrusted.

Most previous works on sensor data aggregation assume a trusted aggregator, and hence cannot protect user privacy against an untrusted aggregator in mobile sensing applications. Several recent works [3] consider the aggregation of time-series data in the presence of an untrusted aggregator. To protect user privacy, they design encryption schemes in which the aggregator can only decrypt the sum of all users' data but nothing else. Rastogi and Nath [3] use threshold Paillier cryptosystem [4] to build such an encryption scheme.

To decrypt the sum, their scheme needs an extra round of interaction between the aggregator and all users in every aggregation period, which means high communication cost and long delay. Moreover, it requires all users to be online until decryption is completed, which may not be practical in many mobile sensing scenarios due to user mobility and the heterogeneity of user connectivity. Different Data Aggregation propose a construction that does not require bidirectional communications between the aggregator and the users, but it has high computation and storage cost to deal with collusions in a large system.

Data aggregation also proposes a construction for sum aggregation, which does not need the extra round of interaction. However, the decryption in their construction needs to traverse the possible plain text space of the aggregated value, which is very expensive for a large system with large plaintext Space.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

Hence, in applications that continuously monitor the carbon dioxide levels that people are exposed to in their daily life [7], the plaintext space can reach 104. Under this plaintext space, for a large system with one million users, the construction in [3] requires 30 seconds to decrypt the sum on a modern 64-bit desktop PC.

Its computation overhead is too high for an aggregator to run real-time monitoring applications with short aggregation intervals and to collect multiple aggregate statistics simultaneously. Moreover, none of these existing schemes considers the Min aggregate (i.e., the minimum value) of time-series data, which is also important in many mobile sensing applications. This paper proposes a new protocol for mobile sensing to obtain the sum aggregate of time-series data in the presence of an untrusted aggregator. The protocol employs an additive homomorphic encryption and a novel key management scheme to ensure that the aggregator can only obtain the sum of all users' data, without knowing individual user's data or intermediate result.

II.RELATED WORKS

2.1. DATA AGGREGATION

Sensor networks composed of small and cost effective sensing devices equipped with wireless radio transceiver for environment monitoring have become feasible. The key advantage of using these small devices to monitor the environment is that it does not require infrastructure such as electric mains for power supply and wired lines for Internet connections to collect data, nor need human interaction while deploying. These sensor nodes can monitor the environment by collecting information from their surroundings, and work cooperatively to send the data to a base station, or sink, for analysis. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offer an increasingly attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity.

Sensor networks are collection of sensor nodes which

Co-operatively send sensed data to base station. As sensor nodes are battery driven, an efficient utilization of power is essential in order to use networks for long duration hence it is needed to reduce data traffic inside sensor networks, reduce amount of data that need to send to base station.

The main goal of data aggregation algorithm is to gather and aggregate data in an energy efficient manner so that Network lifetime is enhanced. Wireless sensor networks (WSN) offer an increasingly Sensor nodes need less power for processing as compared to transmitting data. It is preferable to do in network processing inside network and reduce packet size. One such approach is data aggregation which attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. Wireless sensor networks have limited computational power and limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols. In this proposed system is data aggregation framework on wireless sensor networks is presented. The framework works as a middleware for aggregating data secure measured by a number of nodes within a network.

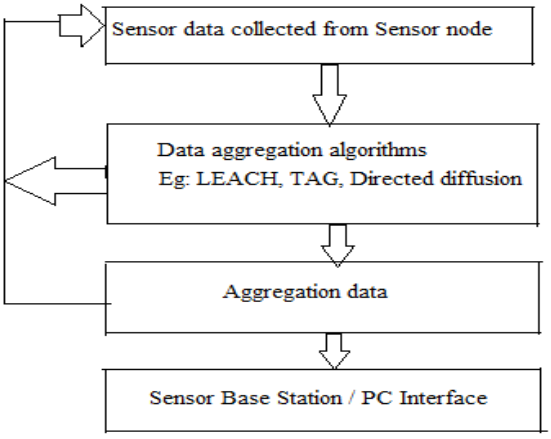


Fig. 1. WSN in Data Aggregation

Fig. 1. Data aggregation is a process of aggregating the sensor data using aggregation approaches. The general data aggregation algorithm works as shown in the below figure. The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH(low energy adaptive clustering hierarchy),TAG (Tiny Aggregation) etc. This aggregated data is transfer to the sink node by selecting the efficient path. There are many types of aggregation techniques are present some of them are listed below.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

2.1.1. CENTRALIZED APPROACH

This is an address centric approach where each node sends data to a central node via the shortest possible route using a multi hop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried. Each intermediate node has to send the data packets addressed to leader from the child nodes. So a large number of messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node.

2.1.2 IN-NETWORK AGGREGATION

In-network aggregation is the global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption, thereby increasing network lifetime. There are two approaches for in-network aggregation: with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining & compressing the data packets received by a node from its neighbors in order to reduce the packet length to be transmitted or forwarded towards sink. In-network aggregation without size reduction refers to the process merging data packets received from different neighbors in to a single data packet but without processing the value of data.

2.1.3. TREE-BASED APPROACH

In the tree-based approach perform aggregation by constructing an aggregation tree, which could be a minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes.

2.1.4 CLUSTER-BASED APPROACH

In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to sink.

III. METHODOLOGY

3.1. SUM AGGREAGATE

In this module, ‘n’, ‘c’ values are given. The product of n*c is calculated. Then the data prepared by the nodes are given as comma separated values. These become x1, x2, … xn. Then key1, key2, … keyn and key0 are calculated. The cipher values are calculated and then cipher sum is created. These values are given to aggregator node.

3.1.1 Protocol Overview

Setup. The key dealer assigns a set of secret values (secrets for short) to each user and the aggregator. Enc. In each time period, user i ($i \in [1,n]$) generates encryption key ki using the secrets that it is assigned. It encrypts its data xi by computing

$$ci = (ki + xi) \bmod M \text{…….(1)}$$

where $M = 2^{\lceil \log2\ (n\ \Delta)\rceil}$. Then, it sends the cipher text ci to aggregator. *AggrDec.* In each time period, the aggregator generates.
Decryption key k0 using the secrets that it is assigned, and decrypts the sum aggregate .

$$S = \sum_{i=1}^{n} xi \text{ by computing}$$

$$S = \left(\sum_{i=1}^{n} ci\text{-}k0\right) \bmod M \text{----------(2)}$$

The keys are generated using a PRF family and a length-matching hash function (see later). According to the aggregator can get the correct sum so long as the following equation holds.

N

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

$k_0 = (\sum_{i=1} ki) \bmod M$----------(3)

Encryption key generation. In time period t ∈ IN, user I generates its encryption key by computing.

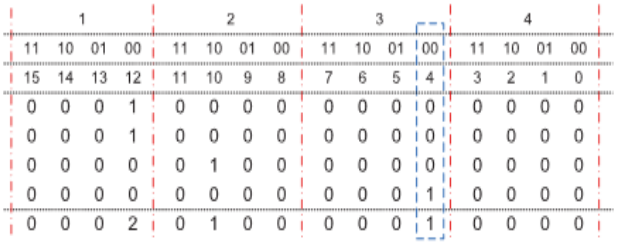
$K_i = (\sum_{s'\in Si} h(fs'(t)) - \sum_{s'\in Si} h(fs'(t))) \bmod M$……(4)

Decryption key generation. Intimeperiod t ∈ IN,the aggregator generates the decryption key by computing

$k_0 = (\sum_{s'\in S} h(fs'(t))) \bmod M$----------(5)

3.2. MIN AGGREAGATE

In this module, 'n', 'c' values are given. The product of n*c is calculated. Then the data prepared by the nodes are given as comma separated values. These become x1, x2, … xn. Then key1, key2, … keyn and key0 are calculated. The cipher values are calculated and then cipher sum is created. These values are given to aggregator node.



1				2				3				4			
11	10	01	00	11	10	01	00	11	10	01	00	11	10	01	00
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	0

Fig.2. Derivative Data[Last Row Is Data Sum of All Users]

In Fig.2. This scheme gets the Min aggregate of each time period using N+1 parallel Sum aggregates in the same time period. The sums used to obtain Min are based on a number of 1 bit derivative data (denoted by d) derived from the users raw data x. Without loss of generality, it is assumed Δ is a power of two. N Aggregate data.

The scheme works as follows: In each time period, each user generates N + 1 derivative data d[0], d[1], ... , d[N], where each derivative data correspond to one possible data value in the plaintext space. For each j ∈ [0, N], the user assigns 1 to d[j] if its raw data value is equal to j and assigns 0 otherwise.

For each j ∈ [0, Δ], the aggregator can obtain the Sum aggregate of d[j] using the sum aggregation protocol presented. Then, Min is the smallest j that returns a positive sum.

In each time period, each user involves in Δ + 1 sum aggregates over Δ + 1 derivative data. Each user uses just one set of secrets for all instances of the sum aggregation protocol.

3.3. MAX AGGREAGATE

In this module, 'n', 'c' values are given. The product of n*c is calculated. Then the data prepared by the nodes are given as comma separated values. These become x1, x2, … xn. Then key1, key2, … key n and key0 are calculated. The cipher values are calculated and then cipher sum is created. These values are given to aggregator node.

3.4. WIRELESS SENSOR NETWORK TREE CONSTRUCTION

In this module, a root node with node id and IP address is keyed in and saved into 'Nodes' table. Then intermediate nodes with corresponding nodes are keyed in with id and parent id. These nodes act as cluster heads for the given group. The child nodes are keyed in with id and parent id (which is a cluster head node id). These nodes are members for the selected group.

3.5. CONCEALED DATA AGGREGATION SCHEME

In this system, **for secure communication between one node and two groups of nodes, CONCEALED DATA AGGREGATION SCHEME (CDAMA) scheme is used**. In the proposed model, a client stores her database on an un trusted service provider. Therefore, the client has to secure their database through PH schemes because PH schemes keep utilizable properties than standard ciphers. Based on PH schemes, the provider can conduct aggregation queries without decryption.

3.6. DATABASE AS A SERVICE MODEL

In DAS model, a client stores her database on an untrusted service provider. Therefore, the client has to secure their database through PH schemes because PH schemes keep utilizable properties than standard ciphers. Based on PH

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

schemes, the provider can conduct aggregation queries without decryption. The most important of all is that we do not have to consider the computation cost and the impact of compromising secret keys (i.e., compromising a client in DAS model is harder than compromising a sensor).

IV RESULTS

PROTOCOL	EXISTING SYSTEM	PROPOSED SYSTEM
Sum Aggregate	Yes	Yes
Min Aggregate	Yes	Yes
Max Aggregate	No	Yes
CDAMA	No	Yes
Database as a Service	No	Yes

TABLE 4.1 PROTOCOLS COMPARISON

Table4.1: This table describes the comparison of existing and proposed system in protocol comparison.

T COMMUNICATION BEHAVIOR	EXISTING SYSTEM	PROPOSED SYSTEM
Group of users count	One	Many
Suitable in Data aggregation.	Wireless sensor network	WSN as well as database (service oriented architecture) environment

TABLE 4.2 COMMUNICATION BEHAVIOR

Table. 4.2 This table describes the difference between existing and proposed system with certain evaluation metrics such as group of users count ,suitable environment.

Scheme	Encryption Field Size	Aggregation (No. of of Computation)	Communication Cost
TinyPEDS	163	2	328
CDAMA (k=2)	768	22	384
CDAMA (k=3)	1024	39	342
CDAMA (k=4)	1280	62	320

TABLE.4.3.ENCRYPTION AND AGGREGATION COST FOR DATA CONCEALMENT SCHEMES
TABLE.4.3This table describes the difference between existing and proposed methods by CDAMA levels.

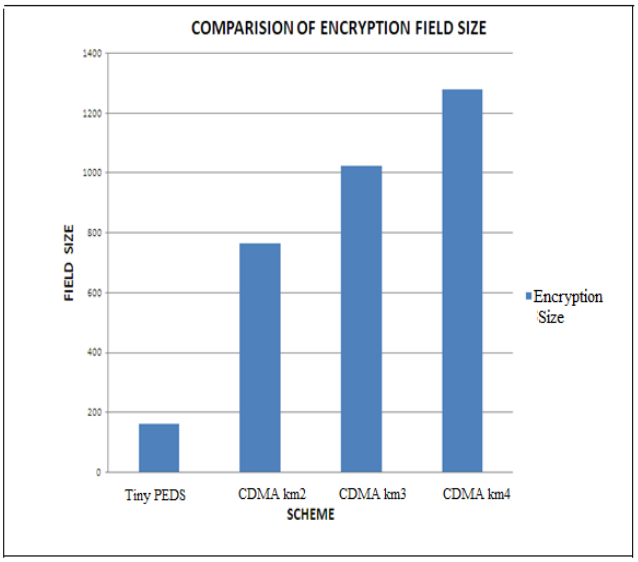


CHART 4.1 COMPARISION OF ENCRYPTION FIELD SIZE

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

This chart describes the difference between existing and proposed methods with certain evaluation metrics such as size of cipher text in bytes.

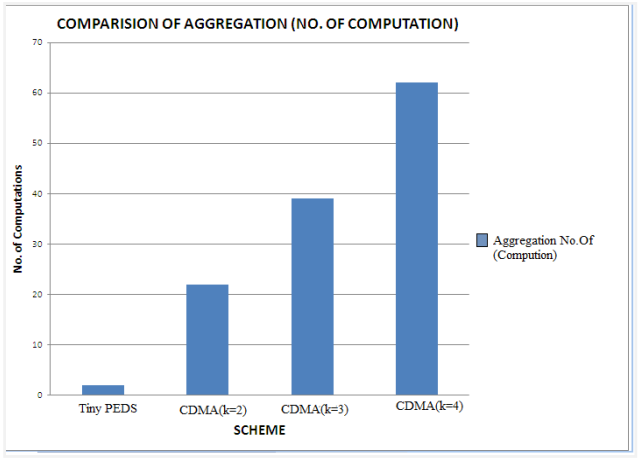


CHART4.2: COMPARISION OF AGGREGATION (NO . OF COMPUTATIONS)

Chart 4.2 This chart describes the difference between existing and proposed methods by Aggregation in CDAMA levels.

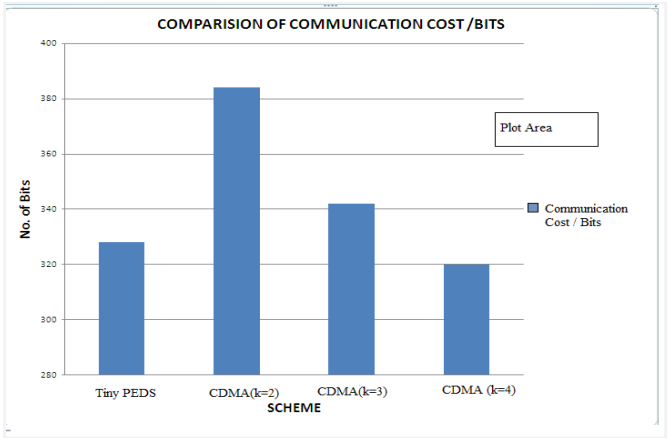


CHART 4.3 COMPARISION OF COMMUNICATION COST/ BITS

Chart.4.3 This chart describes the difference between existing and proposed methods by Communication Bits CDAMA levels.

FINDINGS

- Data aggregation is more secure in Sum, Min, Max Aggregates as well as in CDAMA.
- Through CDAMA, the ciphertexts from distinct applications can be aggregated, but not mixed.
- For a single-application environment, CDAMA is still more secure than other CDA schemes.
- When compromising attacks occur in WSNs, CDAMA mitigates the impact and reduces the damage to an acceptable condition.
- CDAMA is the first CDA scheme that supports secure counting.
- The base station would know the exact number of messages aggregated, making selective or repeated aggregation attacks infeasible.
- The performance evaluation shows that CDAMA is applicable on WSNs while the number of groups or applications is not large.
- Using the Database as a Service Model, the provider can conduct aggregation queries without decryption.

IV.CONCLUSION

This project studied Sum aggregation protocol in WSN environment and it also proposed Min and Max aggregate of time-series data. This project also studied CDAMA scheme for a multi-application environment, which is the first scheme. Through CDAMA, the cipher texts from distinct applications can be aggregated, but not mixed. For a single-

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 10, October 2017

application environment, CDAMA is still more secure than other CDA schemes. When compromising attacks occur in WSNs, CDAMA mitigates the impact and reduces the damage to an acceptable condition. Besides the above applications, CDAMA is the first CDA scheme that supports secure counting. The base station would know the exact number of messages aggregated, making selective or repeated aggregation attacks infeasible. Finally, the performance evaluation shows that CDAMA is applicable on WSNs while the number of groups or applications is not large. In addition, it applied CDAMA to realize aggregation query in Database-As-a-Service (DAS) model. In DAS model, a client stores her database on an untrusted service provider.

REFERENCES

[1] S. Cheng, Z. Cai, J. Li, and X. Fang, "Drawing dominant dataset frombig sensory data in wireless sensor networks," 2015 IEEE Conferenceon Computer Communications (INFOCOM), pp. 531–539, 2015

[2] K. R. Bhakare, R. Krishna, and S. Bhakare, "An energy-efficient grid based clustering topology for a wireless sensor network," International Journal of Computer Applications, vol. 39, no. 14, 2012

[3] M. Shanmukhi and O. Ramanaiah, "Cluster-based comb-needle model for energy-efficient data aggregation in wireless sensor networks," Applications and Innovations in Mobile Computing (AIMoC), pp. 42–47, 2015

[4] Y. Lu, I. Comsa, P. Kuonen, and B. Hirsbrunner, "Dynamic data aggregation protocol based on multiple objective tree in wireless sensornetworks," Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), IEEE, pp. 1–7, 2015.

[5] J. Li, S. Cheng, Y. Li, and Z. Cai, "Approximate holistic aggregation in wireless sensor networks," Proceeding 35th IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 740–741, 2015