International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 03 Issue: 01 | Jan-2016 www.irjet.net p-ISSN: 2395-0072

A SURVEY ON PRIVACY-PRESERVING DATA AGGREGATION WITHOUT SECURE CHANNEL

Kshitija Nandgaonkar[1], Swarupa Kamble[2]

[1] *M.E. Student, Computer Engg ., RMD Sinhgad School of Engineering, Pune ,Maharashtra, India*
[2] *Assistant Professor ,Dept. Computer Engg. ,RMD Sinhgad School of Engineering, Pune, Maharashtra ,India*

Abstract - ***The Privacy-preserving data aggregation problem becoming an important issue in the field of applied cryptography. Most research has been carried out to securely outsource individual's privately owned data to an untrusted aggregator, or to enable multiple parties to jointly aggregate their sensitive data while preserving privacy. However many research require secure pair-wise channel and it suffers from high complexity. This paper describes sum and product calculation protocol that enables an external aggregator or multiple parties to perform data aggregation over participants data while preserving the data privacy.***

Key Words*: Privacy-preserving, Data aggregation, Secure channels, SMC, Homomorphic*.

1. INTRODUCTION

As people are becoming more concerned about their privacy these days, the privacy-preservability is very important. A fundamental problem is that of private data analysis where a third party has to compute some aggregate statistics over some sensitive data held by individuals. This problem finds concrete applications in a number of situations. When the third party, called hereafter aggregator, is trusted an easy solution would be to ask the users to encrypt their data using the aggregator's public key. Upon receiving the ciphertexts the aggregator applies its private key to recover the data in clear and then compute statistics. The problem becomes much more challenging in the case of an untrusted aggregator.

In many real life applications such as crowd sourcing or mobile cloud computing, individuals need to give their delicate data (location-specific or personal information related) to get particular services from the entire system (e.g., location based services or mobile based social networking services).

The data aggregation problem usually involves two different models:

- an external aggregator will gather the data and performs an aggregation function on participants' data (e.g., crowd sourcing);
- participants will together calculate a specific aggregation function where input data being provided by themselves (e.g., social networking services).

However, the individual's data should be kept secret, and the aggregator or other participants are not supposed to learn any useful information about it. Secure Multi-party Computation (SMC), Homomorphic Encryption (HE) and other cryp tographic methodologies can be employed to solve this problem, but these techniques are subject to some limitations in this problem. Many real-world applications have benefitted tremendously from the ability to collect and mine data coming from multiple individuals and organizations. These applications have also spurred numerous concerns over the privacy of user data.

In this paper, we study how an untrusted aggregator or mutiple parties can gather information and learn aggregate statistics over individual privacy. For example, consider a smart grid operator who wishes to track the total electricity consumption of a neighborhood every 15 minutes, for scheduling and optimization purposes. Since such power consumption data can reveal sensitive information about individuals presence and activities, we wish to perform such aggregation in a privacy-preserving manner.

2. LITERATURE SURVEY

C. Castelluccia, A. Chan, E. Mykletun, and G. Tsudik has published paper on Efficient and provably secure aggregation of encrypted data in wireless sensor network [1], in this authors had designed a symmetric key homomorphic encryption scheme which is an addition to homomorphic for conducting the aggregation operations on the ciphertexts. They uses a modular addition, so the scheme is good for CPU bounded devices such as sensor

nodes in WSN. Their scheme can also efficiently compute various statistical values such as mean, variance and deviation. However, since they used the symmetric homomorphic encryption, their aggregator could decrypt each individual sensors data, and they assumed the trusted aggregator in their model.

R. Sheikh, B. Kumar, and D. Mishra, presented a paper on Privacy preserving k secure sum protocol [2], in this paper author proposed a k-secure sum protocol, which is extended from the work of Clifton et al. [3]. They significantly reduced the probability of data leakage occurring in [3] by segmenting the data block of individual party, and distributing segments to other parties. Here, sum of each party's segments is his data, therefore the final sum of all segments are sum of all parties data. This scheme can be easily converted to k-secure product protocol by converting each addition to multiplication. However, pair-wise unique secure communication channels should be given between each pair of users such that only the receiver and the sender know the transmitted segment. Otherwise, each party's secret data can be calculated by performing O(k) computations. In this paper, we remove the limitation of using secure communication channels.

T. Jung, X. Mao, X.-y. Li, S.-J. Tang, W. Gong, and L. Zhang, published a paper on Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation [4], in this paper author assume that all the communication channels in protocol are insecure. Anyone can eavesdrop them to intercept the data being transferred. To address the challenges of insecure communication channel, they assume that the discrete logarithm problem is computationally hard if: 1) the orders of the integer groups are large prime numbers; 2) the involved integer numbers are large numbers. The security of their scheme relies on this assumption. They further assume that there is a secure pseudorandom function (PRF) which can choose a random element from a group such that this element is computationally indistinguishable to uniform random.

M. Joye and B. Libert, has published paper on A scalable scheme for privacy-preserving aggregation of time-series data [5] in this authors have proposed a solution for accommodating large plaintext data from multiple users. In this paper they address problem of private data analysis where a third party has to compute some aggregate statistics over some sensitive data held by individuals. A scheme supports a large plaintext spaces and number of users. It also allows the decryption algorithm to operate in constant time, regardless of the number of users. Additionally the scheme also provides a on-line/off-line efficiency using pre-computations, the encryptor is left with a mere modular multiplication in the on-line phase (i.e., when the data to be encrypted is known), which is highly desirable when computations take place on resource-limited devices. This paper presented a new scheme allowing an untrusted aggregator to evaluate the sum of user's private inputs. In contrast to prior solutions, there is no restriction on the message space or on the number of users. This results in always fast decryption and aggregation, even over large plaintext spaces and/or population of users

We note that Dong et al. [6] investigated verifiable privacy preserving dot production of two vectors and Zhang et al. [7] proposed verifiable multiparty computation, both of which can be partially or fully exploited later. Designing privacy preserving data aggregation while providing verification of the correctness of the provided data is a future work.

Shi et al. [8] proposed a construction that n participants periodically upload encrypted values to an aggregator, and the aggregator computes the sum of those values without learning anything else. This scheme is close to our solution, but they assumed a trusted key dealer in their model. In this paper, the trusted aggregator in [1] is removed since data privacy against the aggregator is also a top concern these days. Unlike [2], we assumed insecure channels, which enabled us to get rid of expensive and vulnerable key pre-distribution. We did not segment each individuals data, our protocols only incur constant communication overhead for each participant. Our scheme is also based on the hardness of the discrete logarithm problem like [8], but we do not trivially employ brute-force manner in decryption, instead, we employ our novel efficient protocols for sum and product calculation.

3. PROBLEM STATEMENT

3.1 Problem Analysis

Assume that there are n participants { p1,p2,....,pn } and each participant pi has a privately known data xi from Zp. The privacy-preserving data aggregation problem is to compute sum or product of xi jointly or by an aggregator while preserving the data privacy. That is, the objective of the aggregator or the participants is to compute the following polynomial without knowing any individual xi:

$$f(x) = \sum_{i=1}^{n} x_i$$

Or

$$f(x) = \prod_{i=1}^{n} x_i$$

Here vector x = (x1; x2,....xn). For simplicity, we assume that the final result f(x) is positive and bounded from above by a large prime number P.

3.2 Security Model

Firstly, we assume that all the communication channels in our protocol are insecure. Anyone can eavesdrop them to intercept the data being transferred. To address the challenges of insecure communication channel, we assume that the following CDH problem is computationally intractable, i.e., any probabilistic polynomial time adversary has negligible chance to solve the following problem:

3.2.1 Definition 1 (CDH Problem in G)
The Computational Diffe-Hellman problem in a multiplicative group G with generator g is defined as follows:
given only g, g^a, g^b ε G where a, b ε Z, compute g^{ab} without knowing a or b.
Additionally, similar Decisional Diffe-Hellman(DDH) problem is defined as follows:

3.2.2 Definition 2 (DDH Problem in G)
The Decisional Diffe-Hellman problem in a multiplicative group G with generator g is defined as follows:
given only g, g^a, g^b , g^c ε G where a, b, c ε Z, decide if $g^{ab} = g^c$.
Our protocol is based on the assumption that it is computational expensive to solve the CDH problem. Then, we define the security of our privacy-preserving sum and product calculation as follows.

3.2.3 Definition 3 (CDH-Security in G)
We say our privacy-preserving (sum or product) calculation is CDH-secure in G if any Probabilistic Polynomial Time Adversary (PPTA) who cannot solve the CDH problem with non-negligible chance has negligible chance to infer any honest participants private value in G.

4. SYSTEM ARCHITECTURE AND DESIGN

There are two models for aggregation of private data as: One Aggregator Model and Participants Only Model. These two models are general cases we are faced with in real applications

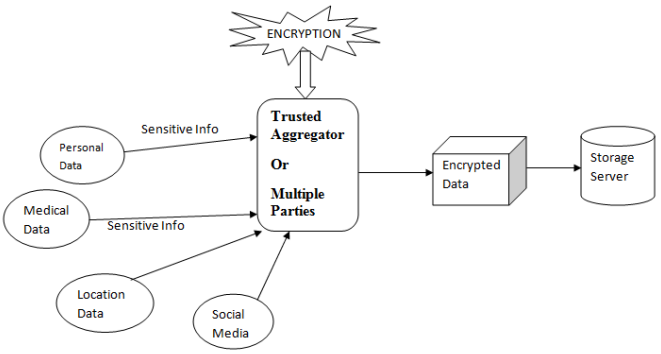

Fig. 1 – System Architecture

A figure shows a system architecture of data aggregation. In this data comes from different user like medical data, location based data, or personal information etc. has been aggregated by external aggregator. An external aggregator or multiple parties will perform encryption on aggregated data to maintain data privacy. The aggregator or participants need to learn privacy preserving sum and product calculation protocol.

4.1 One Aggregator Model

In the first model, we have one aggregator A who wants to compute the function f(x). We assume the aggregator is untrustful and curious. That is, he always eavesdrops the communications between participants and tries to access their input data, but also follow the protocol specification. We also assume participants do not trust each other and that they are curious as well, i.e., they also eavesdrop all the communications and follow the protocol specification. In this model, any single participant pi is not allowed to compute the final result f(x).

4.1.1 Sum and Product Calculation

One Aggregator Model can be used to calculate product and sum of participants privately owned data. In this aggregator A acts as $n+1^{th}$ participant and will compute sum and product for each participants. Here, each participant will broadcast its ciphertext data to aggregator.

4.2 Participant Only Model

The second model is similar to the first one except that there are n participants only and there is no aggregator. In this model, all the participants are equal and they all will calculate the final aggregation result f(x).

4.2.1 Sum and Product Calculation

In participant only model, all participant will jointly compute sum and product of each participants data. In this data is share among all participants of communication. To encrypt data a random key is generated and it is share among all participants.

5. CONCLUSION

This paper presents a privacy preserving sum and product calculation of privately owned data without need of secure channel. Two models can be used for aggregation of data: One aggregator model and Participant only model. A survey describes that, sum and product of sensitive data can be computed without the need of pair wise secure channel. Thus reduces overhead of maintaining each participants key pair.

REFERENCES

[1] C. Castelluccia, A. Chan, E. Mykletun, and G. Tsudik, "Efficient and Provably secure aggregation of encrypted data in wireless sensor networks", *Transactions on sensor Networks(TOSN),* 2009

[2] R. Sheikh, B. Kumar, D. Mishra, " Privacy preserving k-secure Sum Protocol", *Arxiv Preprint arXiv:0912.0956,* 2009

[3] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M.Zhu, " Tools for privacy preserving distributed data mining", *SIGKDD Explorations Newsletter,*2002

[4] T. Jung, X. Mao, X.-y. Li, S.-J. Tang, W. Gong, and L. Zhang, " Privacy-preserving data aggregation without secure channel: multivariate polynomial evaluation," in *INFOCOM.* IEEE, 2013.

[5] M. Joye and B. Libert, "A scalable scheme for privacy-preserving aggregation of time-series data", in *Financial Cryptography and Data Security, (IFCA)* 2013.

[6] W. Dong, V. Dave, L. Qiu, and Y. Zhang, "Secure friend discovery in mobile social networks," in IEEE *INFOCOM,* 2011, pp. 16471655.

[7] L. Zhang, X. Li, Y. Liu, and T. Jung, " Verifiable private multi-party computation: ranging and ranking", in *INFOCOM Mini-conference*, IEEE, 2013

[8] E. Shi, T. Chan, E. Rieffel, R. Chow, and D. Song, " Privacy-preserving aggregation of time-series data", in *NDSS,* vol.17, 2011.