



Preserving Record Sharing Privacy with Anonymous ID Assignments

Prafull B. Masal, Prof. V. S. Kadam

Department of Computer Engineering, Sinhgad Institute Technology,
Lonavala, Pune, Maharashtra, India

Abstract— *Algorithm for anonymous sharing of private data among parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to M. This assignment is anonymous identities in the form of AES and then received are unknown to the other members of the group. Resistance to collusion among other members is verified in an information theoretic sense when private communication channels are used. This assignment of serial numbers allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communications and distributed database access. The required computations are distributed without using a trusted central authority. Existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements. The new algorithms are built on top of a secure sum data mining operation using Newtons identities and Sturms theorem. An algorithm for distribut solution of certain polynomials over finite fields enhances the scalability of the algorithms*

Keywords— *Anonymization and deanonymization*

I. INTRODUCTION

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers remotely store their data the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect privacy of data and oppose unsolicited accesses in the cloud and beyond it, sensitive data, for instance, e-mails, personal health records, photo albums, tax documents, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search[1].

Data mining is a technique that helps to extract useful information from a large database. As the amount of data doubles every three years, data mining is becoming an increasingly important tool to transform this data into information. Data mining tools are increasingly being used to infer trends and patterns. The proposed solution guarantees privacy against most of the attacks known to be possible to retrieve private information of individuals. It also provides the necessary patterns to researchers and data miners without deviating from the original data values. Most importantly the solution does not disturb the distribution of the dataset[2].

The popularity of internet as a communication medium whether for personal or business use depends in part on its support for anonymous communication. Businesses also have legitimate reasons to engage in anonymous communication and avoid the consequences of identity revelation. For example, to allow dissemination of summary data without revealing the identity of the entity the underlying data is associated with, or to protect whistle-blowers right to be anonymous and free from political or economic retributions. Cloud-based website management tools provide capabilities for a server to anonymously capture the visitors web actions. The problem of sharing privately held data so that the individuals who are the subjects of the data cannot be identified has been researched extensively. Researchers have also investigated the relevance

of anonymity and/or privacy in various application domains. patient medical records, electronic voting, e-mail, social networking, etc. Another form of anonymity, as used in secure multiparty computation, allows multiple parties on a network to jointly carry out a global computation that depends on data from each party while the data held by each party remains unknown to the other parties. A secure computation function widely used in the literature is secure sum that allows parties to compute the sum of their individual inputs without disclosing the inputs to one another. This function is popular in data mining applications and also helps characterize the complexities of the secure multiparty computation. This work deals with efficient algorithms for assigning identifiers (IDs) to the nodes of a network in such a way that the IDs are anonymous using a distributed computation with no central authority. Given nodes, this assignment is essentially a permutation of the integers with each ID being known only to the node to which it is assigned. Our main algorithm is

based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data. There are many applications that require dynamic unique IDs for network nodes. Such IDs can be used as part of schemes for sharing/dividing communications bandwidth, data storage, and other resources anonymously and without conflict. The IDs are needed in sensor networks for security or for administrative tasks requiring reliability, such as configuration and monitoring of individual nodes, and download of binary code or data aggregation descriptions to these nodes. An application where IDs need to be anonymous is grid computing where one may seek services without divulging the identity of the service requestor. To differentiate anonymous ID assignment from anonymous communication, consider a situation where parties wish to display their data collectively, but anonymously, in slots on a third party site. The IDs can be used to assign the slots to users, while anonymous communication can allow the parties to conceal their identities from the third party. In another application, it is possible to use secure sum to allow one to opt-out of a computation beforehand on the basis of certain rules in statistical disclosure limitation or during a computation and even to do so in an anonymous manner. However, very little is known with respect to methods allowing agencies to opt-out of a secure computation based on the results of the analysis, should they feel that those results are too informative about their data. The work reported in this paper further explores the connection between sharing secrets in an anonymous manner, distributed secure multiparty computation and anonymous ID assignment. The use of the term anonymous here differs from its meaning in research dealing with symmetry breaking and leader election in anonymous networks. Our network is not anonymous and the participants are identifiable in that they are known to and can be addressed by the others. Methods for assigning and using sets of pseudonyms, have been developed for anonymous communication in mobile networks. The methods developed in these works generally require a trusted administrator, as written, and their end products generally differ from ours in form and/or in statistical properties. To be precise, with nodes the algorithms of this paper distribute a computation among the nodes generating a permutation of chosen with a uniform probability of from the set of all permutations of where will know only. Such a permutation can also be produced by algorithms designed for mental poker. The algorithms for mental poker are more complex and utilize cryptographic methods as players must, in general, be able to prove that they held the winning hand. Throughout this paper, we assume that the participants are semi-honest, also known as passive or honest-but-curious, and execute their required protocols faithfully. Given a semi-honest, reliable, and trusted third party, a permutation can also be created using an anonymous routing protocol. Despite the differences cited, the reader should consult and consider the alternative algorithms mentioned above before implementing the algorithms in this paper. This paper builds an algorithm for sharing simple integer data on top of secure sum. The sharing algorithm will be used at each iteration of the algorithm for anonymous ID assignment (AIDA). This AIDA algorithm, and the variants that we discuss, can require a variable and unbounded number of iterations. Finitely-bounded algorithms for AIDA are discussed. Increasing a parameter in the algorithm will reduce the number of expected rounds. However, our central algorithm requires solving a polynomial with coefficients taken from a finite field of integers modulo a prime. That task restricts the level to which can be practically raised. We show in detail how to obtain the average number of required rounds, and in the Appendix detail a method for solving the polynomial, which can be distributed among the participants.

II. RELATED WORK

Shiba Sampat Kale, Prof. Shivaji R Lahane (1) In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the users data on cloud from the CSP and the third party user. Thus, by hiding the users identity, the confidentiality of users data is maintained.

P.Usha, R.Shriram, W.Aisha Banu (2) It was found that through these experiments only a few attributes in the whole dataset are considered to be sensitive. So the key to privacy preservation is to anonymize these sensitive attributes alone and leave the rest. In this model the same is implemented, by anonymizing the sensitive attributes alone and leaving the rest. Finally the whole dataset to k records was anonymized. The software thus successfully implements the aimed privacy measures without disturbing the privacy as well as the distribution of the dataset.

Ankatha Samuyelu ,Raja Vasanth (3) Main focus is on the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm.

Jain Wang, Yan Zhao , Shuo Jaing, and Jaijin Le (4) The main idea is protecting individuals privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services.

Y. Prasanna, Ramesh (5) This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data.

Larry A. Dunning, Ray Kresman (6) Main objective is to get the access to user data which is stored remotely from anywhere according to user convenience.

J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou (7) Main idea is to formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy.

III. INPUT DESIGN AND OUTPUT DESIGN

A. Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing

can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things.

What data should be given as input?

How the data should be arranged or coded?

The dialog to guide the operating personnel in providing input.

Methods for preparing input validations and steps to follow when error occur.

B. Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

C. Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the systems relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

Convey information about past activities, current status or projections of the Future. Signal important events, opportunities, problems, or warnings. Trigger an action. Confirm an action.

IV. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

A. Interaction Model

1. *Client-driven interventions*: Client-driven interventions are the means to protect customers from unreliable services. For example, services that miss deadlines or do not respond at all for a longer time are replaced by other more reliable services in future discovery operations.

2. *Provider-driven interventions*: Provider-driven interventions are desired and initiated by the service owners to shield themselves from malicious clients. For instance, requests of clients performing a denial of service attack by sending multiple requests in relatively short intervals are blocked (instead of processed) by the service.

B. Modules

1) *Homomorphic encryption Module*: This module to use the first protocol is aimed at suppression-based anonymous databases, and it allows the owner of DB to properly anonymize the tuple t , without gaining any useful knowledge on its contents and without having to send to owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. In order to perform the privacy preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.

2) *Generalization Module*: In this module, the second protocol is aimed at generalization-based anonymous databases, and it relies on a secure set intersection protocol, such as the one found in, to support privacy-preserving updates on a generalization based k -anonymous DB.

3) *Cryptography Module*: In this module, the process of converting ordinary information called plaintext into unintelligible gibberish called cipher text. Decryption is the reverse, in other words, moving from the unintelligible

cipher text back to plaintext. A cipher (or) cypher is a pair of algorithms that create the encryption and the reversing decryption. The detailed operation of a cipher is controlled both by the algorithm and in each instance by a key. This is a secret parameter (ideally known only to the communicants) for a specific message exchange context.

4) *User and Admin Module*: In this module, to arrange the database based on the patient and doctor details and records. The admin to encrypt the patient reports using encryption techniques using suppression and generalization protocols.

5) *Mathematical Representation*:

Set A={a,b,c,d,e,f,...} Is set of System's

Set B={*,^,&,\$,#,@,...} Is set of ID's

Assigning ID's to System's randomly shown below

a----->@ c----->&
 b----->\$ d----->*
 e-----># f----->^

V. SYSTEM ANALYSIS

A. Existing System:

Existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements.. Also, suppose that access to the database is strictly controlled, because data are used for certain experiments that need to be

maintained confidential. Clearly, allowing Alice to directly read the contents of the tuple breaks the privacy of Bob; on the other hand, the confidentiality of the database managed by Alice is violated once Bob has access to the contents of the database. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting Alice and Bob know the contents of the tuple and the database respectively.

Disadvantage:

1. The database with the tuple data does not be maintained confidentially.
2. The existing systems another person to easily access database.

B. Proposed System:

An algorithm for anonymous sharing of private data among parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. Resistance to collusion among other members is verified in an information theoretic sense when private communication channels are used. This assignment of serial numbers allows more complex data to be shared and has applications to other problems

in privacy preserving data mining, collision avoidance in communications and distributed database access. The required computations are distributed without using a trusted central authority.

Advantage:

1. The anonymity of DB is not affected by inserting the records.
2. We provide security proofs and experimental results for both protocols.

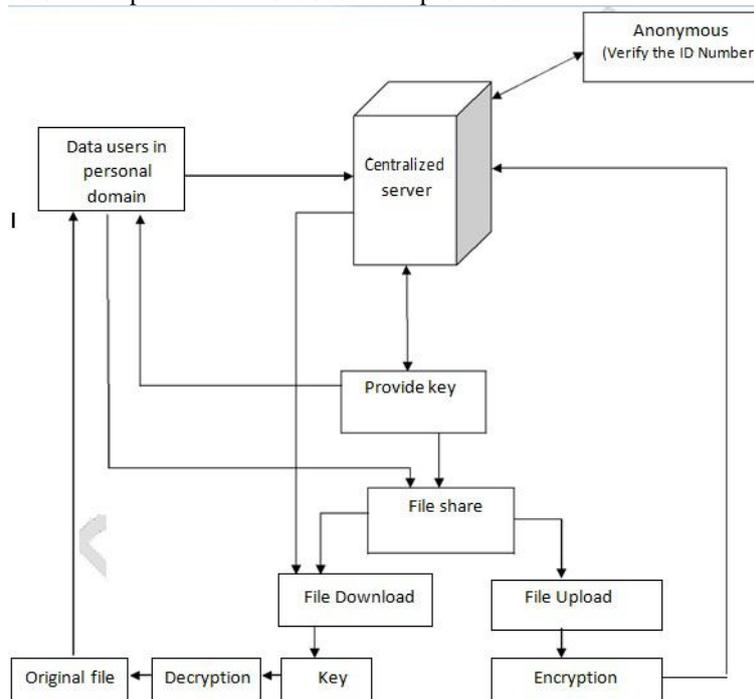
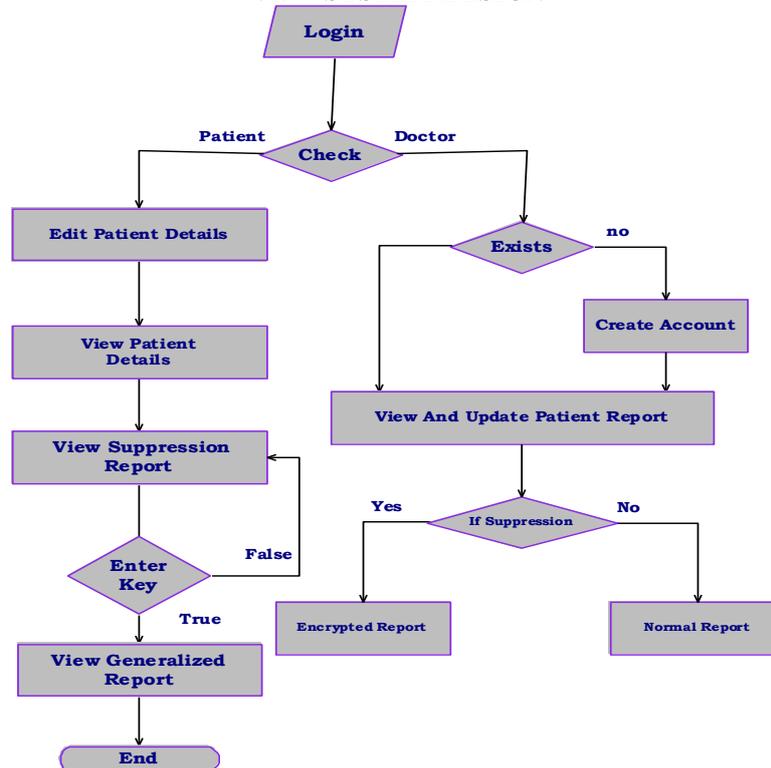


Fig. 1. Proposed system

VI. SYSTEM DESIGN



VII. CONCLUSIONS

In this paper, we studied mainly focused on providing privacy to the data on cloud in which using multikeyword ranked search was provided over encrypted(using AES) cloud data using efficient similarity measure of coordinate matching. The previous work also proposed a basic idea of MRSE using secure inner product computation. There was a need to provide more real privacy which this paper presents. In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the users data on cloud from the CSP and the third party user. Thus, by hiding the users identity, the confidentiality of users data is maintained. Each algorithm compared in Above Section can be reasonably implemented and each has its advantages. Our use of the Newton identities greatly decreases communication overhead. This can enable the use of a larger number of slots with a consequent reduction in the number of rounds required. The solution of a polynomial can be avoided at some expense by using Sturms theorem. The development of a result similar to the Sturms method over a finite field is an enticing possibility. With private communication channels, our algorithms are secure in an information theoretic sense. Apparently, this property is very fragile. The very similar problem of mental poker was shown to have no such solution with two players and three cards. The argument of can easily be extended to, e.g., two sets each of colluding players with a deck of cards rather than our deck of cards. In contrast to bounds on completion time developed in previous works, our formulae give the expected completion time exactly

REFERENCES

- [1] Shiba Sampat Kale, Prof. Shivaji R Lahane, "Privacy Preserving Multi- Keyword Ranked Search with Anonymous ID Assignment over Encrypted Data",2014.
- [2] P.Usha , R.Shriram, "Modified Anonymity Model for Privacy Preserving Data Mining",2013.
- [3] Ankatha Samuyelu Raja Vasanthi, "Secured Multikeyword Ranked Search over Encrypted Cloud Data",.
- [4] Jain Wang, Yan Zhao , Shuo Jaing, and Jaijin Le, "Providing Privacy Preserving in Cloud Computing",.
- [5] Y. Prasanna, Ramesh, "Efficient and Secure Multi Keyword Search on Encrypted Cloud Data",.
- [6] Larry A. Dunning, Ray Kresman, "Privacy Preserving Keyword Searches on Remote Encrypted Data",.
- [7] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing",.
- [8] Ankatha Samuyelu, Raja Vasanthi , "Secured Multi keyword Ranked Search over Encrypted Cloud Data",2011.
- [9] S. Kamara and K. Lauter, "Cryptographic Cloud Storage",Proc. 14th Intl Conf Financial Cryptograpy and Data Security, Jan. 2010.
- [10] A. Friedman, R. Wol, and A. Schuster, "Providing kanonymity in data mining",.