

Data Mining Techniques For Secured Data Sharing And Privacy Preserving Apply Data Mining Approach On Web Service Using AES Algorithm

Author: Snehal Dekate¹; Prof. Jayant Adhikari²; Prof. Sulbha Parate³

Affiliation: Snehal Dekate¹; Prof. Jayant Adhikari²; Prof. Sulbha Parate³

E-mail: dekatesnehal90@gmail.com¹; adhikari.jayant@gmail.com²; sulbha.cse@tgpcet.com³

ABSTRACT

In this paper, the enhancing data techniques are used in the user database for secure their database from other unauthorized user. This technique is useful for privacy preserving; securely share data among N number of parties. And also apply data mining approach on web service.

Keywords: Data Mining, Privacy Preserving, Security, Web Service.

1. INTRODUCTION

1.1 Background of present work:

Algorithms for assigning anonymous IDs are examined with respect to threshold between communication and computational requirements. The new algorithms are built on top of a secure sum data mining operation using Newton's identities and Sturm's theorem. An algorithm for distributed solution of certain polynomials over finite fields enhances the scalability of the algorithms. Markov chain representations are used to find statistics on the number of iterations required, and computer algebra gives closed form results for the completion rates.

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-blower's right to be anonymous and free from political or economic retributions.

Data mining derives its name from the similarities between searching for gold in mines. In gold mines we search for very small particles of gold in tons of soil. Similarly in data mining we search for valuable information from huge amount of data collected in various ways. Data mining, a synonym to "knowledge discovery in databases" is a process of analyzing data from different perspectives and summarizing it into useful information. It is a process that allows users to understand the substance of relationships between data. It reveals patterns and trends that are hidden among the data. It is often viewed as a process of extracting valid, previously unknown, non-trivial and useful information from large databases. Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. If scope of data mining is applied to all events logs generated by various networking devices, system and application servers then efficiency of enterprise security can be drastically increased.

1.2 Scope of present work:

Each algorithm 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 Sturm's theorem. The development of a result similar to the Sturm's method over a finite field is an enticing possibility.

1.3 Objective:

1. Shared secured data.

We can easily share private data among the N parties securely.

2. Preserved privacy.

We can safe privacy of applicant; no one can push the privacy of applicant. That menace there will be no threat to applicant privacy.

3. Apply data mining approach on web service.

If we have to work on any web services, for that we use data mining techniques in large social network.

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.

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 Newton's identities and Sturm's theorem. An algorithm for distributed solution of certain polynomials over finite fields enhances the scalability of the algorithms.

2. Formulation of Present Work

2.1 Dataflow of proposed work:

In proposed algorithm I have to create account for the purpose of use this algorithm. We have account for the authentication from Admin. First of all user create account, and then he can log in. But since admin does not authenticate there account till that he can't sign in his account.

It is useful for secure data .The admin also contain the authenticate ids that's why he share the secure data to authenticate user. All these process done on web services. If admin not authenticate the user then he will goes to stop.

User can upload, download, sharing, mailing, and setting the account information. There are two attackers that is D.DOS And Man in Mediator. In D.DOS if anyone make server busy then in D.DOS have one condition that user can't send continuously more than 5 file to server. In second attackers have Source, Mediator, and Destination. If source send file to destination then mediator can write something on that message and then he will save it, after that he would send message to destination.

2.2 Advanced Encryption Standard:

The Advance Encryption Standard or AES is a symmetric block cipher used by the U.S. government to protect classified information and is implemented in software and hardware throughout the world to encrypt sensitive data.

The origins of AES date back to 1997 when the National Institute of Standards and Technology (NIST) announced that it needed a successor to the aging Data Encryption Standard (DES) which was becoming vulnerable to brute-force attacks.

This new encryption algorithm would be unclassified and had to be "capable of protecting sensitive government information well into the next century." It was to be easy to implement in hardware and software, as well as in restricted environments (for example, in a smart card) and offer good defenses against various attack techniques.

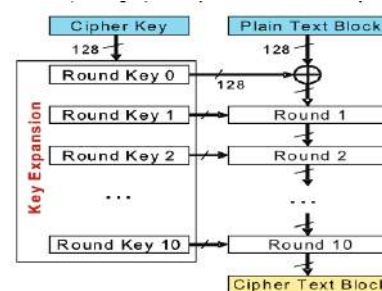


Figure 1. AES Algorithm Structure

Fig.2.1.1 AES Algorithm structure

2.3 Choosing AES

The selection process to find this new encryption algorithm was fully open to public scrutiny and comment; this ensured a thorough, transparent analysis of the designs. Fifteen competing designs were subject to preliminary analysis by the world cryptographic community, including the National Security Agency (NSA). In August 1999, NIST selected five algorithms for more extensive analysis. These were:

- MARS, submitted by a large team from IBM Research RC6, submitted by RSA Security Rijndael, submitted by two Belgian cryptographers, Joan Daemen and Vincent Rijmen Serpent, submitted by Ross Andersen, Eli Biham and Lars Knudsen Twofish, submitted by a large team of researchers including Counterpane's respected cryptographer, Bruce Schneier
- Implementations of all of the above were tested extensively in ANSI, C and Java languages for speed and reliability in encryption and decryption, key and algorithm setup time, and resistance to various attacks, both in hardware- and software-centric systems. Members of the global cryptographic community conducted detailed analyses (including some teams that tried to break their own submissions).

2.4 AES encryption works:

- AES comprises three block ciphers, AES-128, AES-192 and AES-256. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128-, 192- and 256-bits, respectively. (Rijndael was designed to handle additional block sizes and key lengths, but the functionality was not adopted in AES.) Symmetric or secret-key ciphers use the same key for encrypting and decrypting, so both the sender and the receiver must know and use the same secret key. All key lengths are deemed sufficient to protect classified information up to the "Secret" level with "Top Secret" information requiring either 192- or 256-bit key lengths. There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys -- a round consists of several processing steps that include substitution, transposition and mixing of the input plaintext and transform it into the final output of cipher text.

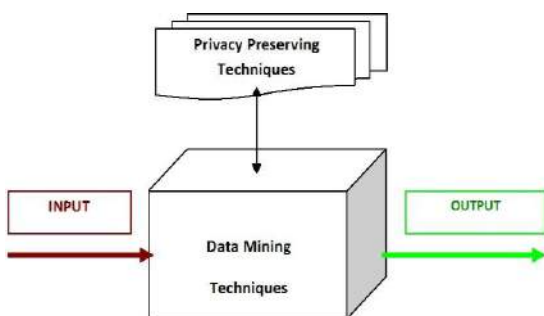


Fig.2.1.2 Input and Output works

2.5 SALIENT FEATURES OF AES:

- AES is a block cipher with a block length of 128 bits. AES allows for three different key lengths: 128, 192, or 256 bits. Encryption consists of 10 rounds of processing for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys. Except for the last round

in each case, all other rounds are identical.

- Each round of processing includes one single-byte based substitution step, a row wise permutation step, a column-wise mixing step, and the addition of the round key. The order in which these four steps are executed is different for encryption and decryption.

3.Result of Analysis:

Encryption algorithm plays very important role in communication security. Our research work surveyed the performance of existing encryption techniques like AES, DES and RSA algorithms. Based on the text files used and the experimental result it was concluded that AES algorithm consumes least encryption and RSA consume longest encryption time. We also observed that Decryption of AES Algorithm is better than other algorithms. From the simulation result, we evaluated that AES algorithm is much better than DES and RSA algorithm. Following table shows that how AES algorithm useful than RSA algorithm.

Algorithm	AES	RSA
Packet size (kb)	153	153
Encryption size	1.6	7.3
Decryption size	1	4.9

Table 3.1 Comparisons of AES and RSA of Encryption and Decryption Time

By analyzing table, Time taken by RSA algorithm for both encryption and decryption process is much higher compare to the time taken by AES algorithm. AES have 128.192,256 bit size of key, and it also have faster hardware and software implementation that's it taken less time for encryption and decryption. It have encryption size 1.6 and RSA have 7.3 that difference shows the RSA algorithm is not much useful than RSA algorithm. Their decryption algorithm also give much time than AES algorithm.

Factors	AES	RSA
Developed	2000	1978
Key size	128.192,256 bits	>1024 bits
Block size	128 bits	Min 512 bits
Ciphering &deciphering key	Same	different

Power consumption	Low	high
Security	Excellent	Least secure
Hardware & software implementation	Faster	Not efficient

Fig 3.2 Comparison between AES and RSA

Above both table shows that AES algorithm is more useful than RSA Algorithm by their different factors like key size , block size, ciphering &deciphering, power consumption, security. AES algorithm have more efficient value than RSA algorithm .the AES algorithm give less time encryption & decryption.

4. Conclusion & Suggested Future Work:

4.1 Conclusion:

In this project, we presented a survey of the broad areas of privacy-preserving data mining and the underlying algorithms. Data modification techniques such as k -Means based techniques. We discussed methods for privacy-preserving mining. Further, we discussed some fundamental limitations of the problem of privacy-preservation in presence of increased amounts of public information and background knowledge. We presented a set of experimental results and also analyzed them from the perspective of data privacy and data utilization. We have also presented a number of diverse application domains for which privacy-preserving data mining methods are useful. Finally, we identified few areas which require further research efforts in the domain of privacy-preserving data mining.

4.2 Suggested Further Work:

Sensitive information of privacy preserving data mining will be present in the forms of analysis information similar to Medicine - hospital cost analysis, prediction hospital cost analysis, drug side effects, and automotive diagnostic expert systems genetic sequence analysis. Finance - credit assessment, fraud detection stock market prediction, Marketing/sales - sales prediction, product analysis, target mailing, identifying unusual behavior, buying patterns, Scientific discovery, Knowledge Acquisition. In addition to that, privacy preserving data mining by implicit function theorem kind of approach will also be used in distributed data mining to protect information of privacy and applied for business

data, which can be represented in the form of vet or valued functions.

5.REFERENCES

- [1] Larry A. Dunning, Member, IEEE, and Ray Kerman "Privacy Preserving Data Sharing With Anonymous ID Assignment" IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 8, NO. 2, FEBRUARY 2013
- [1] Anita Rajendra Zope, Amarsinh Vidhate, and Naresh Harale "Data Mining Approach in Security Information and Event Management" International Journal of Future Computer and Communication, Vol. 2, No. 2, April 2013
- [2] Pasupuleti Rajesh and Gugulothu Narsimha "PRIVACY PRESERVING DATA MINING BY USINGIMPLICIT FUNCTION THEOREM", International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.2, March 2013
- [3] Jianjun Duan, Joe Hurd, Guodong Li, Scott Owens, Konrad Slind, and Junxing Zhang "Functional Correctness Proofs of Encryption Algorithms"
- [4] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, "Online outlier detection in sensor data using non-parametric models, in VLDB", 2006
- [5] Lu-An Tang, Jiawei Han, and Guofei Jiang, "Mining Sensor Data in Cyber-Physical Systems" TSINGHUA SCIENCE AND TECHNOLOGY ISSN 1007-0214 01/11 pp225-234 Volume 19, Number 3, June 2014
- [6] Social Networking Secure Against Malicious Users", 2011 Ninth Annual International Conference on Privacy, Security and Trust.
- [7] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, Michael Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining"
- [8] Dr.R. Sugumar¹, Dr.A. Rengarajan², M.Vijayanand³, "Extending K-Anonymity to Privacy Preserving Data Mining Using Association Rule Hiding Algorithm"