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

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCDSPICE 2016

National Conference on Digital Signal Processing, Information and Communication Engineering

Govt. Polytechnic College, Kasaragod

Vol. 5, Special Issue 4, November 2016

Copyright of Relational Database as a Service (CRDaaS) in secure cloud computing application with Evolutionary and Monte Carlo based Cloud Drop Watermark Approach

Mrs. Divya M B

Lecturer, Dept of Computer Science& Engg, GWPC, Nedupuzha

Abstract: Cloud computing is an emerging technology which has business centric attitude as it is a pay by use model. There is an increasing pace in cloud computi ng business model usage. But a lot of issues are still yet to be solved such as reliability, tolerance, load balancing, security, resource allocation etc. Among them security is the major concern pointed by NIST survey [2]. In cloud environment, user's data has to be released to the cloud and leaves protection sphere of software owner. There are various public key infrastructure based protection schemes available. But these schemes lack trust due to public and private key distribution threats [2]. This paper presents the systematic way of protecting copy right of relational database with a secure design of Software as a Service in cloud environment. Firstly paper explains how public key infrastructure based secure protocol design has to be done with a sample application e-election system. Preliminary research has been done on this problem [1], but there is no comprehensive solution proposed. Hence this paper proposes Monte Carlo estimation and evolutionary based cloud drop watermark approach for database copy right protection. Using secure keys in encrypted form, cloud drops are generated with statistical probabilities for embedding .Monte Carlo estimation on normal probabilities is used for better prediction on mean and standard deviation. Thus this approach eliminates the threat of key distribution. Also this approach maintains the usability of data even after watermarking. Finally, this paper provides comprehensive solution for copy right protection using the above approach, bringing the better randomness and fuzziness through statistical probabilities.

Keywords: Cloud Computing, CRDaaS, Monte Carlo, Public Key, Private Key.

I. INTRODUCTION

Rapid development in the field of internet based business attitude leads to the increased usage of cloud computing. Cloud computing offers economic, scalable benefits to business, but variety of issues are yet to be solved .The major concern is security issue, since user's data has to be released to the cloud and leaves the protection sphere of the owner. [6]

Now, most of the SaaS in the cloud are developed with relational data. Protection of relational data is crucial, so this paper concentrate on it. This paper organized as follows Section II explains the systematic approach for development of SaaS with public key infrastructure based algorithms such as RSA, Blind signature, SHA etc. It explains importance of secure protocol design of SaaS with the help of e-election software. Section III explains the enhanced Monte Carlo based cloud model watermarking for protecting relational data. It also explains crucial digital characteristics such as Mean(M),System Entropy(SE) and System Variance(SV) which are strongly rely on statistical normal (Gaussian) distribution probability. The major factors of this algorithm are Monte Carlo estimation on the existing algorithm [1], improved cryptographic nature of algorithm through evolutionary approach on digital characteristics.

This section also proposes the enhanced watermark cloud model algorithms such as Key generation, Embedding and Extraction algorithms for copy right protection of an RDB in a cloud application.

Section IV analyzes the results of experiments. This paper concludes with summary and direction for future work in section V.

II. SYSTEMATIC APPROACH OF PROTECTING

SaaS (Software as a Service)
Increasing demand of development of cloud applications lead to serious security concerns .While developing an application the following major security steps must be consider.

(1) Develop a secure Protocol for transaction between client and server
(2) Store the sensitive values in database in encrypted form.
(3) Assure the copy right protection of the software.

Firstly all SaaS has to be protected by proper security design with existing cryptographic algorithm such as RSA, SHA, AES etc. In the following section describes a sample application e-voters system [4] that shows how an application has to be designed with proper security constraints .Maximum protection of application has to be done with existing cryptographic algorithm by proper security design.

In e- voters system the major public key based cryptographic algorithms such as RSA digital signature, Blind signature; AES encryption and SHA hashing are used. Even though public key distribution threats remain there, software covering by a strong algorithm improves security.

A. e-VOTERSSYSTEM :SECURE PROTOCOL DESIGN

This is an outline of the scheme for conducting secure elections which is formulated based on standard e-election protocols. The proposed scheme consists of two administrators, The Central Legitimization Agency (CLA) and the Central Tabulating Facility (CTF).

- The CLA keep the voter’s identity and eligibility.
- A voter can log on and cast the vote only after entering AES encrypted password.
- Voter makes **RSA digital signature** on vote and encrypt the vote using **blind signature.** Then both digital signature and blind signed vote is sent to the CLA.
- The CLA on receiving the message verifies the voter’s RSA digital signature. On confirming the eligibility of the voter, a unique pseudo id (**SHA1PRNG**) is generated for the voter. The CLA then signs the blind signed vote with its private key and both the signed vote and the pseudo id is sent back to the voter.
- On receiving the message from the CLA, the voter removes the blinding factor from the vote sends the encrypted vote and the pseudo id to CTF.
- CTF receives the message and sends the id to the CLA for verification. Once the voter is found valid the id is stored in a table by the CTF, to prevent voters from voting more than once. The vote is decrypted and counted.

This application used AES, RSA, Blind signature, SHA1, SHAPRNG. Simiarly any application has to be designed with proper security measures

III. RELATIONAL DATABASE PROTECTION

Monte Carlo based Cloud drop Watermark approach (MCCW)

After the design of SaaS with proper security measures with existing public key based infrastructure algorithms, it is necessary to protect the copyright of data and software. Even cloud model watermarking [1] proposes a solution for copyright of database; this paper makes an improvement over FCG, BCG, SC and watermark embedding and extraction algorithms [1] for getting improved security. The improved watermarking is based on Monte Carlo estimation on Gaussian probability.

Cloud model watermarking is a mathematical probabilistic model used to secure the sensitive data in the database. Unlike the previous cloud drop watermarking approach[1], in MCCW approach cloud drops generate (x,y) co-ordinates in the sample space(numeric attribute field) using mathematical probabilistic theory-Monte Carlo and evolutionary based estimation on Gaussian Probability distribution. Previous approach[1] only concentrating on averaging probabilistic value by considering Gaussian probability, but here uses multiple run on probabilistic values(known as Monte Carlo simulation) to bring more randomness and fuzziness. Generated cloud drops are embedded into the database numeric attribute field .Evolutionary Cloud Drop Watermark Embedding uses evolutionary operator’s mutation and single point crossover, unlike common mathematical operators in previous work [1].

This section covers 3 major algorithms for watermark key generation, embedding and extraction in cloud environment.

A. Evolutionary Monte Carlo Key Generation of Cloud drop algorithm(EMCKGC)

Cloud drops are generated based on major 3 digital parameters such as Mean value (M), System Entropy (SE), System Variance (SV).

Mean Value (M) helps to predict the estimation of some random variable for a long period of trials. From the sensitive attribute field the mean expected value must be calculated and use as M. The expected value of some variable x that takes values x_i can be calculated with n different outcomes, probability of each outcome is p_i is

$$M(x) = x_1p_1 + x_2p_2 + ... + x_np_n$$

In relational database, Arithmetic mean of attribute field can be taken as M such as

$$M = x_1/n+ x_2/n + ... + x_n/n = (x_1 + x_2 + ... + x_n) / n$$

System Entropy means probabilistic distribution function of all elements in the quantitative universe, which used to increase the randomness. Randomness in the embedding

DOI 10.17148/IJARCCE

procedure increases the security. System Entropy can be defined as

$$SE=-k\sum P_i \ln(p_i)$$

Where K= R/ $N_{avagadro}$
R= Universal gas constant (8.3143 J/Mol)
$N_{avagadro}$ =6.02*10^{23} Molecules/mole
P_i = Precise description of randomness (In relational database, it may be a single quantity which is a function of P_i).

System variance is entropy of entropy which is also called hyper Entropy in probabilistic theory. It can be defined as standard error of mean such as
$S_e = S_x / \sqrt{\mathbf{N}}$
S_x =Standarad deviation
N=Number of cases.
Monte Carlo based cloud drop generation algorithm in detail is

Input: Number of tuples in RDB table(Nt :Integer)
Numeric attribute field value(Val :Array)

Output: Returns (x,y) cloud drop co-ordinates for embedding.

Algorithm:
(1) Calculate Mean (M)

$$M=\frac{\sum_{i=1}^{Nt} \text{Val [i]}}{Nt}$$

(2) Calculate System Entropy(SE)

$$SE= -k \sum_{i=1}^{Nt} Pi\ln(Pi)$$

Condition that SE>=3

(3) Calculate System Variance(SV)
SV=$S_x / \sqrt{\mathbf{Nt}}$

$$S_x=\sqrt{\sum_{i=1}^{Nt}(\text{Val}[i]-M)^2/Nt}$$

(4) Store initial values of M,SE,SV in table along with particular user's RDB.Do the steps 5,6 for finding evolutionary random values on SE,SV and M to make more cryptic randomness
(5) Store the initial integer part of digital parameters into digital array.
(6) Do Evolutionary digital characteristics generation and obtain different values on each run and store the sequence in array named run.
(7) Select run[(i+1)%3]
(8) Generate random cloud drops (X,Y) using Monte Carlo estimation such as
X=Mean[Normal Distribution(M,SV)]
Y=Mean[Normal Distribution(M,SE
(9) Repeat steps 7-9 until Nt cloud drops are generated

A.1 EVOLUTIONARY DIGITAL CHARACTERISTICS GENERATION

Three digital characteristics values are generated using basic equations. Then create a tree using following algorithm to produce different digital parameters on each run.

```
Function CreateEvolutioaryTree(digital parameter :array):run array

(1) Take integer values of digital parameters
(2) For each digital parameter
(3)     swap(digitalparmeter[i],
            digitalparmeter[(i+1)%3],)
(4)     store swaped digital parameter into
            run array.
(5)         for each element in swap array
(6)             diiference[i]←difference[i]+run[i]-
                run[i+1]
(7)Sort difference array along with run array
(8) Return run array
```

The major strengths in this algorithm compared to Forward Cloud drop Generator (FCG) [1] are

- FCG algorithm [1] produces a constant probabilistic digital value as it has only single run. In each numeric attribute field, this constant digital parameter value is used. So by analyzing multiple RDB's, it is easy to hack. But in MCCW approach provides secrecy by producing different values on each run .And also in previous work [1] uses simple addition operation but this approach doesn't uses normal mathematical operators to obtain different digital parameters run values, rather producing it by Hill climbing tree generation.
- The major advantage of EMCKGC algorithm is that it removes the key distribution problem. The generated initial keys are kept in administrator database (Cloud provider) .Valid user's has to contact with cloud provider for extraction of watermark.
- Unlike FCG, MCCW algorithm uses RDB's number of tuples and numeric attribute fields for producing three digital parameters. Since each RDB varies in these parameters, algorithm produces more randomness and fuzziness.
- FCG uses only Normal probability distribution which produces values with different nature,so in MCCW, averaging the behavior of cloud drop by Monte Carlo estimation. (i.e. Mean value of Gaussian distribution). Monte Carlo estimation produces unique random number generation on long period run,and store it in cloud drop table.

B. Monte Carlo Based Cloud drop Watermark Embedding Algorithm (MCCWE)

Monte Carlo based keys are used for encryption. This algorithm uses various evolutionary operations such as mutation and crossover. These operators help to bring more fuzziness into the embedding procedure.

Input: Numeric attribute field value (Val []: integer)
Number of tuples (Nt: Integer)
Clouddrops(x,y:co-ordinates)

Output :MCCWE

Algorithm:

```
(1) Convert A. Val[i] into bit array
(2) For i←1 to Nt
(3) crossval[i]←SPCROSSOVER(A.val[i])
(4) Do MUTATION(crossval[i])
(5) Convert X[i] into bit array
(6) result[i]  ←crossval[i]⊕ X[i]
(7) Convert Y[i] into bit array
(8) while RDB.Eof
(9)       Select numeric attribute field value A.val[i]
(10)            if A.val[i] is not null then
(11)                  A.val[i]←result[i]⊕Y[i]
(12)                  Store integer value of A.val[i]
(13)      i←i+1
(14)      RDB.next
(15)      End
```

The major strength in this embedding procedure is as follows

- Advantage of using Single Point Crossover (SPCROSSOVER) rather than normal mathematical operators provides the better hiding of sensitive information and cloud users can't sense the watermark.
- Even though watermark is embedded with evolutionary operators, usability of protected data to authorized users is not limiting.
- Even though algorithm uses the MUTATION operator, which actually makes inversion of origin (no relation to parent data), still integrity of data will not be lost.
- Compared to Cloud water mark embedding algorithm [1], no particular scaling factor required for this algorithm since scaling of embedding done with the proper evolutionary stages.

C. Cloud provider's algorithm for Cloud drop Watermark Extraction (CCWEx)

Protected database can be accessed by extracting the watermark. But water mark extraction must be controlled by the cloud provider, since the watermark information (key and cloud drop information) is only stored at cloud providers database.

Input: Initial key values M, SE, SV
Cloud drop table, X, Y
RDB attribute value(A.val[]:integer)

Output: Watermark extracted RDB

Algorithm:

```
(1) For i←1 to Nt
(2) Apply Evolutionary digital characteristics generator
    approach of initial M,SE,SV and store the generated
    sequence in key array
(3) For i←1 to Nt
(4) M=key[i].First,SE=key[i].Next,SV=key[i].Next
(5) Check for all X[i] is equal to
    Mean[Gaussian(M,SV)],if not go to step 13
(6) Check for all Y[i] is equal to
    Mean[Gaussian(M,SV)],if not go to step 13
(7) For i←1 to Nt
(8)     result[i]←A.val[i]⊕Y[i]
(9)     mut[i]←result[i]⊕X[i]
(10)        MUTATION(mut[i])
(11)        SPCROSSOVER(mut[i])
(12)      Store the cross over result into RDB
(13)     Alert cloud provider that malicious database
    detected.
(14)     END
```

BCG and SC algorithm [1] checks only whether the RDB is malicious or not .But CCWEx algorithm used by the cloud provider to prevent the malicious software services (SaaS) from the cloud environment. The valid initial digital parameters are kept by the cloud provider in encrypted form(AES or any approach) .Only the valid user, verified by the cloud provider ,has right to change the watermark. If a developer needs to change his watermark, he has to contact with the cloud provider. Cloud provider will delete the existing watermark and after that developer can add a new watermark on it. Watermark on already marked RDB is not allowed by the cloud provider.

An application developer who wishes to make his services available on cloud has to contact with the cloud provider. The software is available on the cloud only after watermark is embedded by the cloud provider using CCWEx algorithm.

IV. EXPERIMENT RESULT AND ANALYSIS

Proposed method validated by running sample data sets on the following configuration. Configuration is virtualization environment using VirtualBox,version 4.0.8,My SQL server version with 5.5 ,Windows XP professional version with 2.10GHZ CPU,4GB RAM, and 500GB hard disk. The programming language used is java(JDK1.6) using the tool Net Beans version 6.8 .Various datasets regarding consumer details and rating are downloaded and loaded into My SQL database server for cloud users to be available.

Datasets are varying in number of tuples from 100 to 3000 tuples. Embeded cloud drops are the major encryption component in the watermarking. The central theme of any of watermarking method is to protect the copyright of data or application without affecting the content. So regarding RDB, data must be usable to the cloud users even after watermarking .The cloud drop value rate must be minimum enough to maintain this rule. Compare to the method [1] ,in which cloud drop value ranges between -10 to10 ,but MCCW approach reduces value range to -1 to 4 .The implication is that Monte Carlo estimation and evolutionary strategies on cloud model algorithm improves the efficiency rate of watermarking in terms of usability even data watermarked. Cloud drop generation chart developed from sample dataset through java JFree chart library depicted in figure 2.

Cloud drops value fall within the range of -1 to 4,irrespective of number of tuples.The cirlcle plots on the graph indicate the cloud drop points with 1557 tuples and rectangular points indicates the cloud drop points with 2000 tuples.Thus the generation of cloud drops indicates the improved efficiency of algorithm in terms of usability even after watermarking.

Cloud model watermarking[1] proposes a model which can be used by the cloud provider only if original RDB is with them.The adverse effect of that approach is the memory wastage of keeping the entire database along with watermark information.But in MCCW approach cloud provider has to keep only watermark information,there is no need of entire original RDB.Thus this approach not only increases the usability but also improves the memory efficiency.As the watermark is unique on gaussian random method with smaller cloud drop ranges it is so hard to delete the watermark from the RDB.Thus the copy right of RDB data improves through this approach.

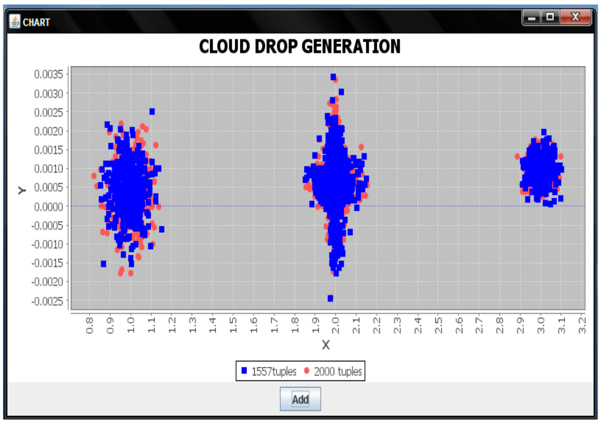

Figure 2: Cloud drop generation

The main idea behind watermark is that it won’t distort the original data. As in this approach Monte Carlo estimation creates only positive values on smaller Gaussian probability due to the sample set generates only positive values.

In figure 3, depicts the right side curve of Gaussian curve which obtains due to Monte Carlo estimation on normal probability on smaller range values. Most of the RDB tuples are numerical positive values .By avoiding the negative Gaussian values with Monte Carlo estimation, less watermark distortion will happen.

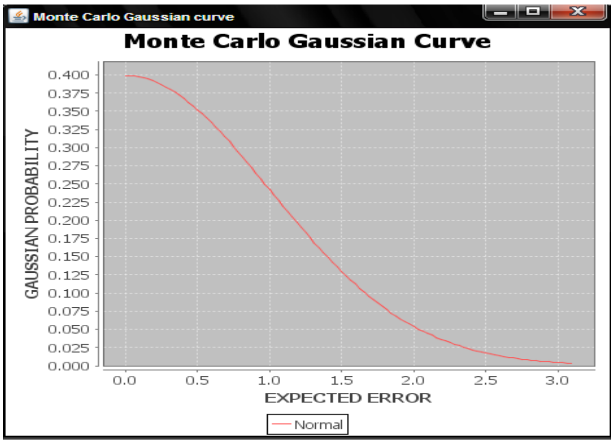

Figure 3: Monte Carlo Gaussian curve on positive value

In figure 4 ,Even though the tuples contain negative values ,curve become similar to Normal probability distribution curve like as method proposed by cloud model watermarking[1].But range of value reduces to -1 to 4 from -10 to 10.So MCCW approach produces better result.

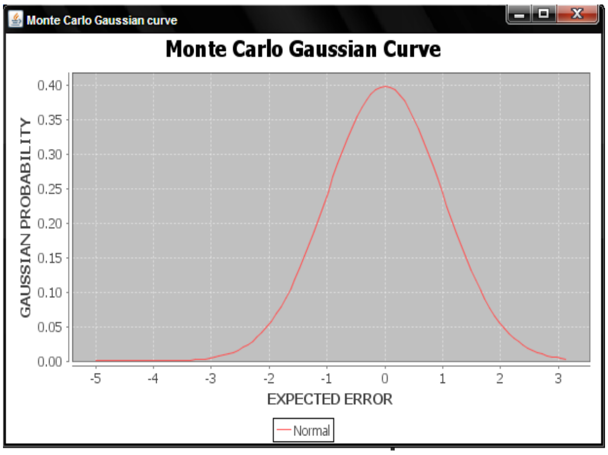

Figure 4: Normal Monte Carlo Gaussian curve

V. FUTURE WORK AND CONCLUSION

Thus the major threat of distribution of private key in cryptographic algorithms in cloud environment can be solved by cloud model watermarking. The better randomness and fuzziness are offered in this proposed MCCW approach rather than cloud model [1] algorithms. Thus an application can be secured with proper security design protocols like e-voter’s system and copyright of database can be established through evolutionary and Monte Carlo based watermark approach and can provide it as a service.

This paper focuses on Cloud model watermarking on copy right protection of RDB since data is more sensitive information in modern world. The future trend in this approach is to make use of MCCW approach for software copyright. It can be done by name protection or jar file protection or any static or dynamic watermarking on code can be established with proper designing of algorithm. Similarly other future trend is to use this cloud model watermarking on BLOB (Binary Large Object) data rather than other data types, since it provides more security.

ACKNOWLEDGMENT

Mrs. Divya M.B thanks to head of the institution, colleagues and her family members for their support.

REFERENCES

[1] Yong Zhang, Xiamu NIU and Dongning Zhao,"A method of protecting relational databases copy right with cloud watermark" ,International journal of Information technology, Volume[1],No[3], IEEE Computer Society

[2] Zhiwei Yu, Chaokun Wang and Jianmin Wang," A Novel Watermarking Method for Software Protection in the Cloud", Software practice and experience, Volume [00], No [1-23], Wiley InterScience, November 2010

[3] Janakiram MSV, Cloud computing strategist, "Demystifying the cloud: An introduction to cloud computing" Volume 1.0, March 2010.

[4] Janga Sireesha, So-In Chakchai, Secure Virtual Election Booth with Two Central Facilities, Department of Computer Science Washington University in St. Louis, USA, December 14th 2005

[5] Kai Hwang and Deyi Li Tsinghua, "Trusted Cloud Computing with Secure Resources and Data Coloring", IEEE INTERNET COMPUTING, Volume [1089], No [801/10], IEEE Computer Society

[6] Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing, April 2009.

[7] Blind Signature [Online] http://www.rsa.com/rsalabs/node.asp?id=2339

[8] SHA1 Encryption Algorithm [Online] http://www.vocal.com/data_sheets/SHA1.pdf

[9] Statistical Mechanics Entropy, Order Parameters, and Complexity James P. Sethna Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY

BIOGRAPHY

Mrs. Divya M B, Now Lecturer in Govt Women's Polytechnic College, Nedupuzha. She completed B. Tech in Computer Science & Engineering from Calicut University in 2006.From 2007-2015; she worked as a Lecturer in Sahrdaya college of Engineering. She completed her ME in computer science & engineering from PSG College of engineering. She handled the subject areas such as compilers, theory of computation, data structures and analysis, graph theory and combinatorics. Her area of interest is data structures and computational issues.