IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-4, Issue-6 E-ISSN: 2347-2693

Preserving Privacy using Column Masking and Data Encryption Techniques

Lakshmi B[1*], Ravindra Babu H[2] and Murali Krishna A[3]

[1*] Department of Computer Applications, V.R.S.E.C, Vijayawada -7, Andhra Pradesh, India
[2] IT company, Bangalore, India
[3] Department of Electronics and Communications, RVR & JC College of Engineering, Andhra Pradesh, India

www.ijcseonline.org

Received: May/16/2016 Revised: May/30/2016 Accepted: Jun/18/2016 Published: Jun/30/ 2016

Abstract—Information Technology plays an important role in Multi-National corporations with thousands of computer systems to small business organizations with a single system. Every business organization maintains IT Sector which is charged with maintaining and monitoring information. As rapid increase in information technology and omnipresent of internet has caused information breaches to grow faster. An information breach is a security incident in which confidential data of the organization has been stolen and used by unauthorized personnel. Information breaches may involve financial, personal information such as credit card or bank details, personal identification details, health insurance details and trade secret information of organizations. So protecting such confidential data has become a vital and integral part of each and every organization. This paper deals with security breaches, vulnerabilities in information systems and demonstrates security with column masking and data encryption techniques.

Keywords- *Data Masking, Column Masking and Data Encryption*

I. INTRODUCTION

IT sector maintains large volumes of database and plays a prominent role in providing security and protects the confidential information from security breaches. Information technology emerging into new trends to secure information, intruders' is using much more advanced techniques to break the security.

Most of the organizations agreed that their staff were involved in some of the information breaches and they suffered with the problem of loss of data. Protecting information from intruders' is as important as providing secured logins and passwords. It really helps in controlling the breaches of security laws and regulations of an organization.

II. SECURITY BREACHES

Security policies and procedures are violated to access confidential data is called security breach or security violation. Intruders bypass the underlying security mechanisms of Business applications, services, and networks to access unauthorized logical IT Parameters.

Every organization maintains different type of hardware and software firewall to detect the violation of security regulations. It notifies the issues to security administrator.

Security Breach is an event that compromises the confidentiality, integrity, or availability of data. It's less severe than an Information breach. Information breach is a confirmed disclosure of data to an unauthorized party. This is more serious than a security breach.

According to 2015 Cost of Data Breach analysis by Ponemon Institute, This year, 47% of all breaches were found to be the result of a malicious or criminal attack.

The three main reasons of increased cost for data breaches are

- Data breaches become more frequent and the costs of remediating the repercussion have also increased.
- The lost business resulting from a data breach has also increased, representing the damage to company reputation which customers are increasingly aware of thanks to a number of high ,profile breaches.
- Customers are increasingly aware to a number of high profile breaches results in damage to the company's reputation leads to loss of business.
- Increased cost of breach detection and escalation.

The following graph shows broken down by volume of records to the likelihood of a breach occurring. The probability of a breach of 10,000 records occurring is 25% and chance of a breach of 100,000 or more records occurring is less than 1%.

*Corresponding Author:
B. Lakshmi
e-mail: itslakshmi.h@gmail.com

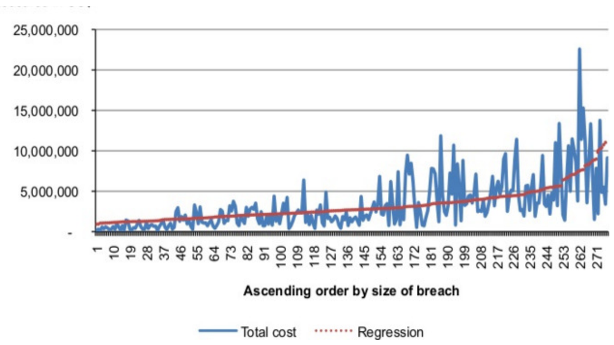


Figure 1: Graph between likelihood of breach and volume of records

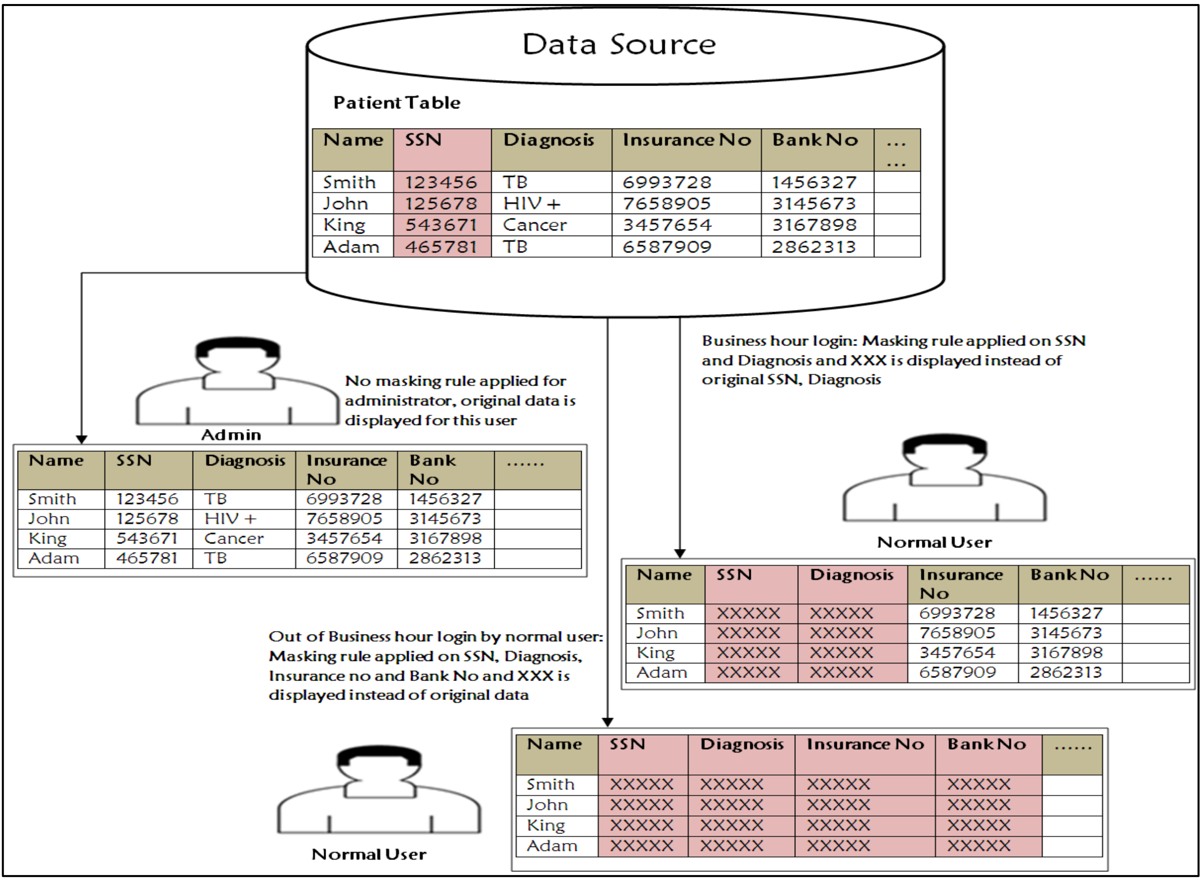
III. COLUMN MASKING

The process of hiding valuable data from certain users without having to apply encrypt/decrypt techniques is called data masking. Data masking also called data scrambling or data anonymization

The following types of users participate in the data masking process for a typical enterprise:

- Database administrator of Application or application developer: This user has complete knowledgeable about the application and database objects.
- Information security administrator: This user defines information security policies, enforces security best practices, and also recommends the data to be hidden and protected.

Column masking is new database concept that is to address the shortcomings of column access control methods. Column masks are applied based on the accessibility of user on the particular column. For example, a column mask defined on the phone number column of patient table ensures that Doctor sees only phone numbers of the patients who accept to share their phone numbers with him. For other patients, the phone number would be set to NULL or masked out based on column mask definition.



Patient Table

Name	SSN	Diagnosis	Insurance No	Bank No	… …
Smith	123456	TB	6993728	1456327	
John	125678	HIV +	7658905	3145673	
King	543671	Cancer	3457654	3167898	
Adam	465781	TB	6587909	2862313	

No masking rule applied for administrator, original data is displayed for this user

Name	SSN	Diagnosis	Insurance No	Bank No	......
Smith	123456	TB	6993728	1456327	
John	125678	HIV +	7658905	3145673	
King	543671	Cancer	3457654	3167898	
Adam	465781	TB	6587909	2862313	

Business hour login: Masking rule applied on SSN and Diagnosis and XXX is displayed instead of original SSN, Diagnosis

Name	SSN	Diagnosis	Insurance No	Bank No	......
Smith	XXXXX	XXXXX	6993728	1456327	
John	XXXXX	XXXXX	7658905	3145673	
King	XXXXX	XXXXX	3457654	3167898	
Adam	XXXXX	XXXXX	6587909	2862313	

Out of Business hour login by normal user: Masking rule applied on SSN, Diagnosis, Insurance no and Bank No and XXX is displayed instead of original data

Name	SSN	Diagnosis	Insurance No	Bank No	......
Smith	XXXXX	XXXXX	XXXXX	XXXXX	
John	XXXXX	XXXXX	XXXXX	XXXXX	
King	XXXXX	XXXXX	XXXXX	XXXXX	
Adam	XXXXX	XXXXX	XXXXX	XXXXX	

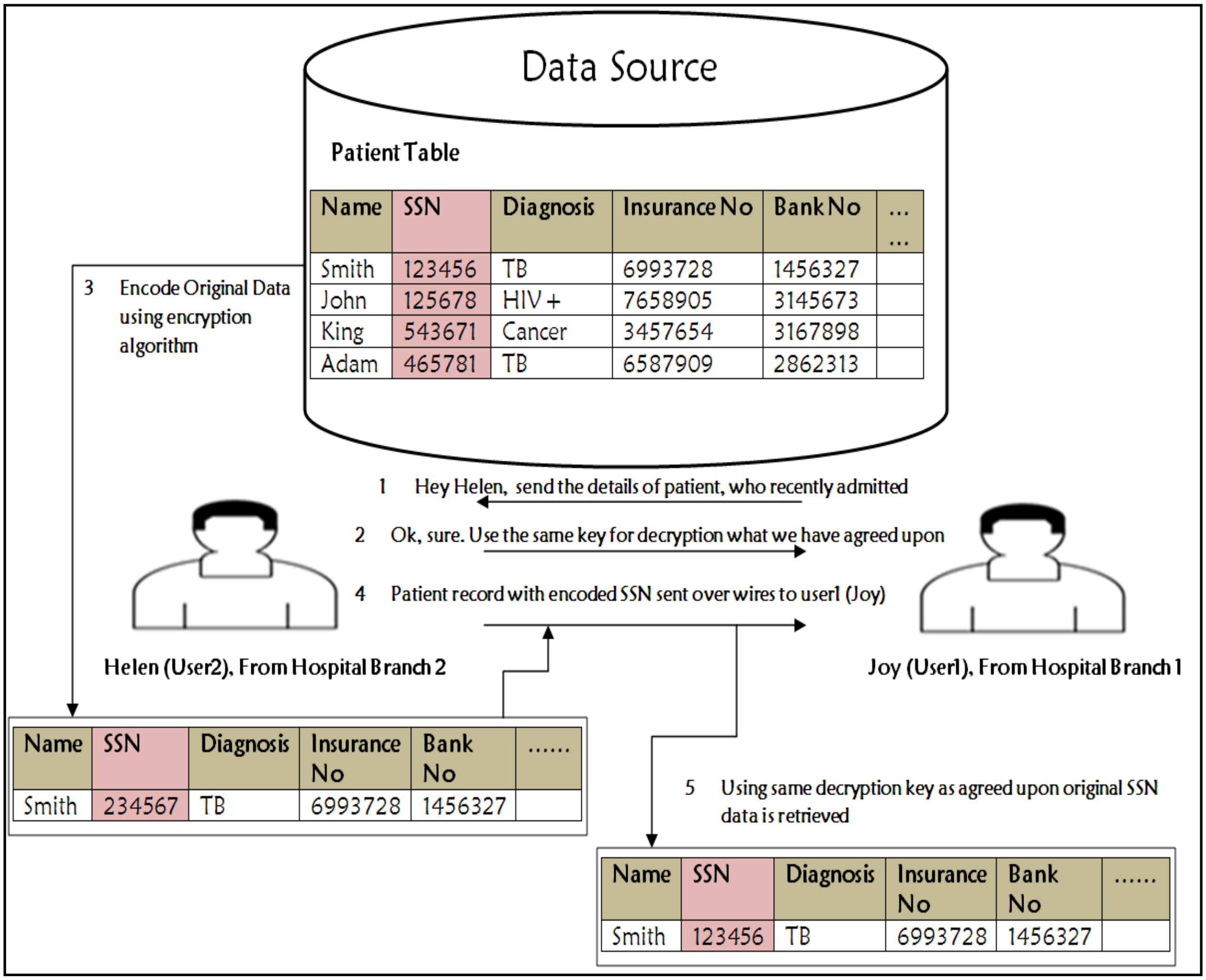
Figure 2: Data masking for sensitive information like SSN, Bank account number, Insurance Number

The three major advantages of column masking are:

- No database user can exempted from column masking, not even users with Database Access authority. Only the user with database security administrator authority has the ability to manage column mask definitions.

- Table data is protected despite of hoe table is accessed such as through ad hoc query tools, an application, or through report generation tools.

- Column access control is transparent to existing applications. Changes in applications are doesn't take any advantage of them. Column masks no longer checks what is being asked but rather who is asking what. That is, based on the context in which the query was asked the result set changed, no warnings or errors are returned.


Data Source
Patient Table
3 Encode Original Data using encryption algorithm
1 Hey Helen, send the details of patient, who recently admitted
2 Ok, sure. Use the same key for decryption what we have agreed upon
4 Patient record with encoded SSN sent over wires to user1 (Joy)
Helen (User2), From Hospital Branch 2
Joy (User1), From Hospital Branch 1
5 Using same decryption key as agreed upon original SSN data is retrieved

Name	SSN	Diagnosis	Insurance No	Bank No	… …
Smith	123456	TB	6993728	1456327	
John	125678	HIV +	7658905	3145673	
King	543671	Cancer	3457654	3167898	
Adam	465781	TB	6587909	2862313	

Name	SSN	Diagnosis	Insurance No	Bank No	……
Smith	234567	TB	6993728	1456327	

Name	SSN	Diagnosis	Insurance No	Bank No	……
Smith	123456	TB	6993728	1456327	

Figure 3: Encryption and decryption cycle using symmetric encryption algorithm

IV. DATA ENCRYPTION

The process of translating data into a secret code is called data encryption. Data Encryption is one the most effective ways of securing confidentiality of digital data stored on computer systems and data transmitted using the Internet or other computer networks. To decrypt data, user must have access to a secret key or password that used to encrypt it. Unencrypted data is called plain text; encrypted data is referred to as cipher text.

Advantages of Data Encryption:

- Encryption separates the security of data from the security of the devices where the data resides or the medium through which data is transmitted.
- Implementing encryption provides the provision of strong protection for intellectual property.
- System or users are authorized to read the data. i.e., system or user can only read the encrypt data only when they has the key to decrypt the data
- Encryption protects both data at rest and data in flight. Like two sides of a coin Encryption also has **flaws**.

Disadvantages of Data Encryption:

- Data Encryption is a very complex technology. Encryption keys Management is an administrative task which often overburdened IT staff.
- Data encryption completely relates to keys, now the security of information becomes security of the encryption key.
- Lose of Encryption key means effectively lose of data.
- Encrypting data and creating the encryption keys and securing the keys are computationally expensive.
- No matter what type of encryption is used, the systems performing the computational heavy lifting must have available resources.
- A poor encryption implementation could result in a false sense of security, when in fact it is wide open to attack.

Example:

Let us consider a patient database.

- A user1 (Joy) from one branch requests the patient details who had admitted in another branch from Helen, who is authorized to access the data source.
- They both agreed to use the encryption/decryption key.
- Helen encodes the sensitive data, like SSN, by incrementing each digit by 1and sends it to Joy.
- Joy knows the decryption key, so she decodes the SSN and gets original data.

V. METHODS OF ENCRYPTION

There are three basic encryption methods:

- Hashing
- Symmetric Cryptography
- Asymmetric Cryptography

Each of these encryption methods has their own uses, pros, and cons. For example, Hashing is very resistant to tampering, but is not as flexible as the other methods.

A. *Hashing Encryption*

The first encryption method, called hashing, creates a unique, fixed-length signature for a message or data set. Hashes are created with an algorithm, or hash function, and people commonly use them to compare sets of data. Since a hash is unique to a specific message, even minor changes to that message result in a dramatically different hash, thereby alerting a user to potential tampering.

A key difference between hashing and the other two encryption methods is that once the data is encrypted, the process cannot be reversed or deciphered. This means that even if a potential attacker were able to obtain a hash, he or she would not be able to use a decryption method to discover the contents of the original message. Some common hashing algorithms are Message Digest 5 (MD5) and Secure Hashing Algorithm (SHA).

B. *Symmetric Methods*

Symmetric cryptography, also called private-key cryptography, is one of the oldest and most secure encryption methods. The term "private key" comes from the fact that the key used to encrypt and decrypt data must remain secure because anyone with access to it can read the coded messages. A sender encodes a message into cipher text using a key, and the receiver uses the same key to decode it.

People can use this encryption method as either a "stream" cipher or a "block" cipher, depending on the amount of data being encrypted or decrypted at a time. A stream cipher encrypts data one character at a time as it is sent or received, while a block cipher processes fixed chunks of data. Common symmetric encryption algorithms include Data Encryption Standard (DES), Advanced Encryption Standard (AES), and International Data Encryption Algorithm (IDEA).

C. *Asymmetric Forms*

Asymmetric or public key, cryptography is, potentially, more secure than symmetric methods of encryption. This type of cryptography uses two keys, a "private" key and a "public key," to perform encryption and decryption. The use of two keys overcomes a major weakness in symmetric key cryptography, since a single key does not need to be securely managed among multiple users.

In asymmetric cryptography, a public key is freely available to everyone and used to encrypt messages before sending them. A different, private key remains with the receiver of ciphertext messages, who uses it to decrypt them. Algorithms that use public key encryption methods include RSA and Diffie-Hellman.

VI. MASKING VERSUS ENCRYPTION

There are many similarities between both technologies, though the differences are substantial. Each of them is designed to ensure data protection, which can be substantially improved when both are used in synergy.

- Encryption is great for securing data that needs to be returned to its original value at some point – e.g. during transmission, sitting in a production database, etc.

- Encryption is not a suitable replacement for masking when data would not need to be returned to the original value – e.g. development, unit & QA testing, etc.

- Encryption secures the data, but only while the encryption keys are safe. If the encryption key is discovered, it can unlock everything. With most data masking strategies, there is no "master key". When it comes to ensuring non-production data is truly safe and never compromised, no matter what happens to it.

- Encryption poses technical issues if it is being used to secure production data in non-production environments. For example, a small name like John would be a very long string of characters when encrypted. This means that database tables might not be able to store the entire encrypted string. Same goes for applications using that data, UI designs might not be able to accommodate strings that large.

- If encryption is used to secure a database or specific tables, generally, the application using that data needs to decrypt it before showing it on the screen. This means that although the data is protected from users hitting the database to browse the tables, sensitive information would still be available in the application using that database. This is obviously desired for production scenarios, however, in development/QA this is a security risk.

- Data masking solves the risk issues associated with data in non-production environments and non-production activities by permanently changing the data in a consistent, repeatable fashion using the original data, so you end up with data that is production quality, but without the production risk.

A. *Encryption:*

Original data is encoded (Intermediate encrypted data) and from this encoded data the original data is retrieved.

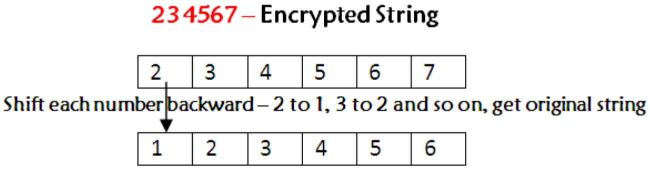

234567 – Encrypted String
2 3 4 5 6 7
Shift each number backward – 2 to 1, 3 to 2 and so on, get original string
1 2 3 4 5 6

Figure 4: Encryption

B. *Masking:*

Original data is masked / obscured and there should not be any provision to retrieve the original data from the obscured data

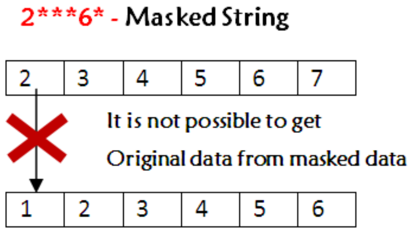


Figure 5: masking

Data masking does not encrypt information. User can see all data records in its native form and no decryption key is necessary. But user will see only what user is allowed to see today and not a byte more. And tomorrow user may see even less, if the rules will change overnight.

Best ciphers can be cracked (may be in a million years using today's technology), while masked data cannot be unmasked. Resulting data set does not contain any references to the original information. That makes it absolutely useless for the attackers.

CONCLUSION

It is easy to think of data masking and data encryption as the same things, since they are both data-centric means of protecting sensitive data. However, it is their inherent procedures and purposes that differentiate them. Both Data Masking and Data Encryption are relatively easy to implement. As long as user knows what to perform and how to implement these technologies, user can easily protect both productive and non-productive information.

Preserving the data privacy will be achieved up to some extent by using these two techniques. But still there is a threat when users are getting increased and most of them are acting as administrators. It is happening because of that they may be familiar with encryption algorithms or else they may be the password holders.

REFERENCES

[1] B. Lakshmi, K. Parish Venkata Kumar, A. Shahnaz Banu and K. Anji Reddy, "Data Confidentiality and Loss Prevention using Virtual Private Database.", International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 5 No. 03 Mar 2013 143.

[2] Ravikumar G K, Manjunath T N, Ravindra S Hegadi, Umesh I M, "A Survey on Recent Trends, Process and Development in Data Masking for Testing" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

[3] Ajayi, Olusola Olajide, Adebiyi, Termidayo Olarewaju, " Application of Data Masking in Achieving Information Privacy" IOSR Journal of Engineering, ISSN (e): 2250-3021, ISSN (p): 2278-8719, Vol. 04, Issue 02 (February. 2014), ||V1|| PP 13-21.

[4] Insiders Behaving Badly, IEEE Security & Privacy (Volume:6 , Issue: 4)

[5] A Review on Database Security, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[6] http://spdp.di.unimi.it/papers/wiley.pdf

[7] http://www.cse.hcmut.edu.vn/~c503002/Files/TRUONG QuynhChi/Slides/Chap9_DBS.pdf

[8] https://www.dtc.umn.edu/umssia/resources/day7a_08.pdf

[9] Emil Burtescu, "Database Security - Attacks And Control Methods", Journal of Applied Quantative Methods http://www.jaqm.ro/issues/volume-4,issue-4/pdfs/burtescu.pdf

Authors Profile

Mrs. Lakshmi B is currently working as Asst. Professor for Department of Computer Applications at Velagapudi RamaKrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh. She has 10 years of teaching experience. Her areas of interest include Privacy Preserving Data Mining Techniques, Computer Networks and Big Data Analytics. She had received MCA from Acharya Nagarjuna University and M.Tech from JNTUK, Kakinada. She had ratified under Nagarjuna and JNTUK, Kakinada. She had completed the OCA certification.

Mr. Ravindra Babu H is presently working as Associate Manager in an IT company, Bangalore. He has 12+ years of IT working experience in JAVA/J2EE/Webservices technologies. His areas of interest include Data structures, and Big Data. He completed his M.Tech in C.S.E from VIT University Vellore. He is currently pursuing MBA from SMU.

Mr. Murali Krishna A is presently working as Asst. Professor for the Department of Electronics and Communications at RVR & JC College of Engineering, Chowdavaram, Guntur, Andhra Pradesh.