DATA SECURITY

Suraj.K.H
Computer Science and Engineering
RNS Institute of Technology

Shashank.M
Computer Science and Engineering
RNS Institute of Technology

ABSTRACT:

Databases are a collection of organized information that can easily be accessed, managed and updated.Database systems plays a vital role in corporate world because they communicate information related to your sales transactions, product inventory, customer profiles and marketing activities..Database security can guard against a compromise of your database, which can lead to financial loss, reputation damage, consumer confidence disintegration, brand erosion, and non-compliance of government and industry regulation.

INTRODUCTION:

Data security includes every aspect of information security from the physical security of
hardware and storage devices to administrative and access controls, as well as the logical security of software applications. It also includes organizational policies and procedures.
When properly implemented, robust data security strategies will protect an organization's information assets against cybercriminal activities, but they also guard against insider threats and human error, which remains among the leading causes of data breaches today.

Data Security

Data security involves deploying tools and technologies that enhance the organization's visibility into where its critical data resides and how it is used. Ideally, these tools should be able to apply protections like encryption, data masking, and redaction of sensitive files, and should automate reporting to streamline audits and adhering to regulatory requirements.

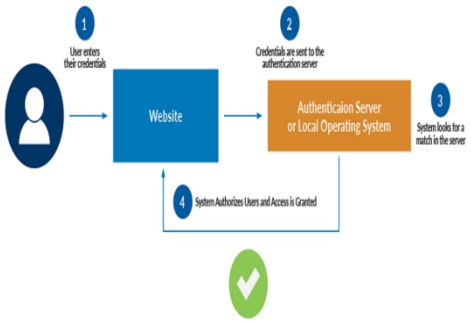
Examples of data security technologies include backups, data masking and data erasure. A key data security technology measure is encryption, where digital data, software/hardware, and hard drives are encrypted and therefore rendered unreadable to unauthorized users and hackers.One of the most commonly encountered methods of practicing data security is the use of authentication. With authentication, users must provide a password, code, biometric data, or some other form of data to verify identity before access to a system or data is granted.The importance of data security will help us to formulate a general plan to protect that data. There are many data security technologies and processes that can support to increase the company's productivity and its efficiency in protecting the confidential data of the users.As we all know that in today's world privacy is given and considered as a top most priority of every individual all over the globe.

Authentication:

Authentication, along with authorization, is one of the recommended ways to boost data security and protect against data breaches. Authentication technology verifies if a user's credentials match those stored in your database. Today's standard authentication

processes include using a combination of ways to identify an authorized user, such as passwords, PINS, security tokens, a swipe card, or biometrics.

Authentication is made easier through single sign-on technology, which, with one security token, allows an authenticated user access to multiple systems, platforms, and applications. Authorization technology determines what an authenticated user are allowed to do or see on your website or server.



Aunthentication Process between the user and the server

Let us consider the scenario where a user enters their credentials in the website then:

1.Authetication server needs to know exactly who is accessing their information from the respective site.
2.Authentication is used by a client when the client needs to know whether the system meets it's expectations that is discussed while buliding the website for the customer .

3.During Authentication phase, the user has to prove his identity to the server for further process,so that the server can track iwhether he has a authorised access to the site.

4.Authentication by a server entails the use of a user name and password. Other ways to authenticate can be through Cards, Retina scans, Voice recognition, and Fingerprints.
5.Authentication does not determine what tasks the individual can do or what files the individual can see. Authentication just identifies and verifies who the person or system.

Whenever the user tries to login with the valid crenditials such as user name and password then Authentication and authorization happen through the process called access control. Access control systems can include the following activities:

- Discretionary access control-which allows access to resources based on the identity of users or groups.
- Role-based access control,-which assigns access based on organizational role and allows users access only to specific information.
- And mandatory access control-which allows a system administrator to strictly control access to all information.

Backups and Recovery:

While prioritizing data security also requires a plan for how to access your company's and client's data in the event of system failure, disaster, data corruption, or breach. Doing regular data backups is an important activity to help with that access.

A data backup is responsible for making a copy of your data and storing it on a separate system or medium such as a tape, disk, or in the cloud. You can then recover lost data by using your backup.

Back Up and Restore Relationship

SMS Backup and Restore

Mannual Backup :is manually downloading and creating backups for all your files and data.

Automatic Backup:is basically enabled through backup software that automates the entire backup process. Typically, automatic backup first requires an administrator to configure the systems/network that need to be backed up.The administrator just needs to specify the type and time of the automatic backup within the backup software.

Repository Export:With this functionality we can easily export data from one database to another rather than creating new data in different databasesLets us consider SQL which is a structure query language which manages data through relational database management system.Wecan easily export data from one database to another using
the startSQLRepository utility.

Encryption:

Data encryption software effectively enhances data security by using an algorithm (called a cipher) and an encryption key to turn normal text into encrypted ciphertext. To an unauthorized person, the cipher data will be unreadable.

Data encryption translates data into another form, or code, so that only people with access to a secret key (formally called a decryption key) or password can read it. Encrypted data is commonly referred to as ciphertext, while unencrypted data is called plaintext. Currently, encryption is one of the most popular and effective data security methods used by organizations. Two main types of data encryption exist - asymmetric encryption, also known as public-key encryption, and symmetric encryption.

That data can then be decrypted only by a user with an authorized key. Encryption is used to protect the data that you store (called data at rest) and data exchanged between databases, mobile devices, and the cloud (called data in transit). Your encryption keys must be securely managed, including protecting your critical management systems, managing a secure, off-site encryption backup, and restricting access.

Symmentric Encrptyion:

Symmentric Encrpytion a type of encryption where only one key (a secret key) is used to both encrypt and decrypt electronic information. The entities communicating via symmetric encryption must exchange the key so that it can be used in the decryption process. This encryption method differs from asymmetric encryption where a pair of keys, one public and one private, is used to encrypt and decrypt messages.

By using symmetric encryption algorithms, data is converted to a form that cannot be understood by anyone who does not possess the secret key to decrypt it. Once the intended recipient who possesses the key has the message, the algorithm reverses its action so that the message is returned to its original and understandable form. The secret key that the sender and recipient both use could be a specific password/code or it can be random string of letters or numbers that have been generated by a secure random number generator (RNG). For banking-grade encryption, the symmetric keys must be created using an Random Number Generator that is certified according to industry standards.

There are two types of symmetric encryption algorithms:

Block algorithms. Set lengths of bits are encrypted in blocks of electronic data with the use of a specific secret key. As the data is

being encrypted, the system holds the data in its memory as it waits for complete blocks.

Stream algorithms. Data is encrypted as it streams instead of being retained in the system's memory.

We will consider the DES cipher block Algorithm here:

DES is a block cipher, and encrypts data in blocks of size of 64 bit each, means 64 bits of plain text goes as the input to DES, which produces 64 bits of cipher text. The same algorithm and key are used for encryption and decryption, with minor differences. The key length is 56 bits.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64

Figure - discording of every 8th bit of original key

Discarding the 8th bit of the key

We are basically, discarding of every 8th bit of the key produces a 56-bit key from the original 64-bit key.
The discarded bits are the multiples of 8 that is-8,16,24,32,40,48,56,64

DES is based on the two fundamental attributes of cryptography: substitution (also called as confusion) and transposition (also called as diffusion).

DES consists of 16 steps, each of which is called as a round. Each round performs the steps of substitution and transposition. Let us now discuss the broad-level steps in DES.

The Block Cipher Algorithm with respect to DES(Data Encryption Standards) is as followed:
In the first step, the 64 bit plain text block is handed over to an initial Permutation (IP) function.
The initial permutation performed on plain text.

Next the initial permutation (IP) produces two halves of the permuted block; says Left Plain Text (LPT) and Right Plain Text (RPT).

Now each LPT and RPT to go through 16 rounds of encryption process.
In the end, LPT and RPT are rejoined and a Final Permutation (FP) is performed on the combined block
The result of this process produces 64 bit cipher text.

Initial Permutation (IP) –

As we have noted, the Initial permutation (IP) happens only once and it happens before the first round. It suggests how the transposition in IP should proceed, as show in figure.

For example, it says that the IP replaces the first bit of the original plain text block with the 58th bit of the original plain text, the second bit with the 50th bit of the original plain text block and so on.

This is nothing but jugglery of bit positions of the original plain text block. the same rule applies for all the other bit positions which shows in the figure.

As we have noted after IP done, the resulting 64-bit permuted text block is divided into two half blocks. Each half block consists of 32 bits, and each of the 16 rounds, in turn, consists of the broad level steps outlined in figure.

Step-1: **Key transformation –**

We have noted initial 64-bit key is transformed into a 56-bit key by discarding every 8th bit of the initial key. Thus, for each a 56-bit key is available.
From this 56-bit key, a different 48-bit Sub Key is generated during each round using a process called as key transformation. For this the 56 bit key is divided into two halves, each of 28 bits.

These halves are circularly shifted left by one or two positions, depending on the round.

For example, if the round number 1, 2, 9 or 16 the shift is done by only position for other rounds, the circular shift is done by two positions. The number of key bits shifted per round is show in figure.

Round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
#key bits shifted	1	1	2	2	2	2	2	2	1	2	2	2	2	2	2	1

Figure - number of key bits shifted per round

Key Bits Shifiting Process

After an appropriate shift, 48 of the 56 bit are selected. for selecting 48 of the 56 bits the table show in figure given below.

Bit number 18 is discarded (we will not find it in the table), like 7 others, to reduce a 56-bit key to a 48-bit key. Since the key transformation process involves permutation as well as selection of a 48-bit sub set of the original 56-bit key it is called Compression Permutation.

Now we have a table of 48 bit keys,and each time a different subsets of key bits is used in each round which inturns leads to difficulting in cracking the code as they are unique in nature.

14	17	11	24	1	5	3	28	15	6	21	10
23	19	12	4	26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	40	51	45	33	48
44	49	39	56	34	53	46	42	50	36	29	32

Figure - compression permutation

Compression Permutation

Because of this compression permutation technique, a different subset of key bits is used in each round. That's make DES not easy to crack.

Step-2: **Expansion Permutation** –

Recall that after initial permutation, we had two 32-bit plain text areas called as Left Plain Text(LPT) and Right Plain Text(RPT).
During the expansion permutation, the RPT is expanded from 32 bits to 48 bits. Bits are permuted as well hence called as expansion permutation. This happens as the 32 bit RPT is divided into 8 blocks, with each block consisting of 4 bits.

Then, each 4 bit block of the previous step is then expanded to a corresponding 6 bit block, i.e., per 4 bit block, 2 more bits are added.

Here in case of RPT we have 32 bits,by considering 8bits in each block,we land up with 4 blocks with 8 bits eachso to form a 48 bit value we need to add 2 bits to each of the blocks which results with 6 bits in each blocks,and eventually makes up a total of 48 bits

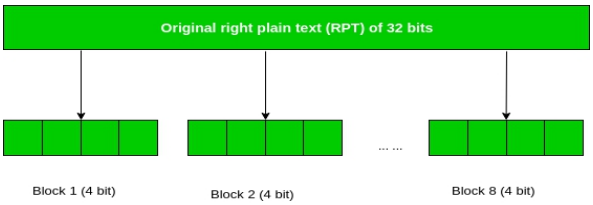


Figure - division of 32 bit RPT into 8 bit blocks

Division of 32 bit RPT(Right Plain Text)

This process results into expansion as well as permutation of the input bit while creating output.Key transformation process compresses the 56-bit key to 48 bits. Then the expansion permutation process expands the 32-bit RPT to 48-bits.

Now the 48-bit key is XOR with 48-bit RPT and resulting output is given to the next step, which is the S-Box substitution.

Asymentric Encryption:

Asymmetric Encryption uses two distinct related keys. One key, the Public Key, is used for encryption, and the other, the Private Key, is for decryption. As implied in the name, the Private Key is intended to be private so that only the authenticated recipient can decrypt the message.

Here private key has to remain private which is used to decrpty the information of the message which is recieved from the public key that is nothing but the encrypted message.

Both the keys are mathematically connected with each other. This relationship between the keys differs from one algorithm to another.

The algorithm is basically a combination of two functions – encryption function and decryption function. To state the obvious, the encryption function encrypts the data and decryption function decrypts it.

Asymmentric Encryption Algorithm:

Cryptographic hashing algorithm, also known as hash functions, basically scramble data. A hash function will generally take an arbitrary amount of data, apply a mathematical formula, and produce a fixed length product, called the hash value.

Sometimes, you will also hear the original data referred to as the message, and the product is referred to as the message digest. Hashing is mostly used as a secure way of storing data.

Hashing relies on a couple of key principles. The first is the fact that hashes are one-way; that is, you can use the hash and the data to create the hash value, but you cannot figure out the data given the hash value. Hash functions should also avoid collisions.

A collision is where two different sets of data produce the same hash value. Third, you should not be able to change data without having the hash value also change.

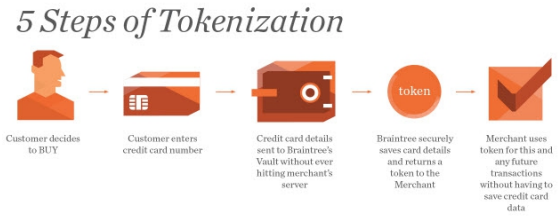
Tokenization

Tokenization is used to substitute sensitive data with random characters that are not algorithmically reversible.

The relationship between the data and its token values is stored in a protected database lookup table, rather than being generated by and decrypted by a mathematical algorithm (as in the case of encryption).

The token representing the real data is used across different systems as a replacement, while the actual data is stored on a separate, secure.

Types of tokens:

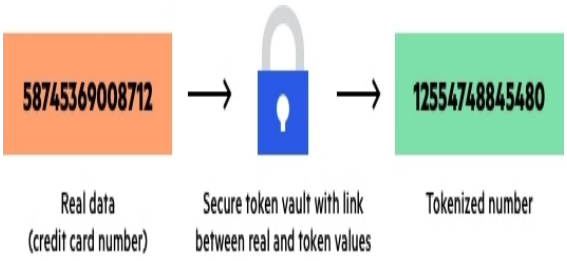


Tokenization

There are many ways that tokens can be classified however there is currently no unified classification. Tokens can be: single or multi-use, cryptographic or non-cryptographic, reversible or irreversible, authenticable or non-authenticable, and various combinations thereof.In the context of payments, the difference between high and low value tokens plays a significant role.

Broadly considering the different types of tokens it is mainly divided into two parts:

1. High Level Tokens
2. Low Level Tokens



High Value Tokens

High-value tokens (HVTs)

HVTs serve as surrogates for actual PAN(Primary Account Number) in payment transactions and are used as an instrument for completing a payment transaction. In order to function, they must look like actual PANs.

Multiple HVTs can map back to a single PAN and a single physical credit card without the owner being aware of it.

Additionally, HVTs can be limited to certain networks and/or merchants whereas PANs cannot.

HVTs can also be bound to specific devices so that anomalies between token use, physical devices, and geographic locations can be flagged as potentially fraudulent.

Ex: When a merchant processes the credit card of a customer, the PAN is substituted with a token. 1234-4321-8765-5678 is replaced with, for example, 6f7%gf38hfUa.

(The Above process is Known as Tokenization)

The merchant can apply the token ID to retain records of the customer, for example, 6f7%gf38hfUa is connected to Mary. The token is then transferred to the payment processor who de-tokenizes the ID and confirms the payment.
6f7%gf38hfUa becomes 1234-4321-8765-5678.

(Now the de tokenization takes Place)
Due to tokenization the customer data(Primary Account Number) recieved by the merchant is confedential and is protected from various cyber threats and the transaction takes places very smoothly by maintaining the customer satisfaction.

Low-value tokens (LVTs)

Low-value tokens (LVTs) also act as surrogates for actual PANs in payment transactions. LVTs cannot be used in and of themselves to complete a payment transaction.
For LVTs to work at all, it must be possible to match them back to the actual PANs they represent.

A consumer's PAN is tokenized by replacing the actual value with the surrogate value, the token. The token must always be matched back to the actual PAN to complete a payment transaction.

This mapping from LVT to actual PAN is done within a "tokenization system."

Ex:In this scenario,the token(6f7%gf38hfUa) should match with the customer PAN(1234-4321-8765-5678) ,only then the transaction takes place between the merchant and the Customer.

Referrences:

1. https://www.forcepoint.com/cyber-edu/data-security
2. https://searchsecurity.techtarget.com/definition/authentication
3. https://en.wikipedia.org/wiki/Tokenization_(data_security)
4. https://www.tutorialspoint.com/cryptography/data_encryption_standard.htm

Authors:

Suraj.K.H-currently pusruing B-Tech degree majoring Computer Science and Engineering From RNS Institute of Technology Banglore and even has interests in Cryptography,Cipher Algorithms and Cryptocurrency.

Shashank M-currently pursuing B-Tech degree B-Tech degree majoring Computer Science and Engineering From RNS Institute of Technology Banglore and even has interests in Cryptography,Cryptocurrency,Cipher Algorithms.