



SESSION 4

# Cryptography - 2

## Learning Objectives

- Describe cryptography
- Explain symmetric key cryptography
- Explain asymmetric key cryptography
- Describe digital signatures
- Explain cryptographic hashing
- Explain the difference between hashing and encryption
- Discuss how to use public keys as identities

## Introduction

Every organization has confidential data which is accessible only to the authorized users. This data can include client details, emails with critical information, file servers with important information, websites with important client information, etc. Organizations have an obligation to keep this data secure and encryption allows organizations to use various tools to accomplish this task. Cryptography is applied to a variety of fields where the integrity of data is of paramount importance. Blockchain, e-commerce and fintech sector are some of the important areas where cryptography has been employed.

Cryptography has become all the more relevant in light of the security breaches that have become frequent across all the major organizations around the globe.

The main objective of cryptography is to provide protection by transforming data into an unreadable form. The word "cryptography" comes from the Greek words "kryptós", which means "hidden", and "graphein", which means "write". Cryptography employs different techniques that provide security across a network, verify the authenticity of data, lock the files, act as the delivery proof of sent messages, and so on.

In this session, you will learn the basics of cryptography. After this, you will learn about some of the most common symmetric and asymmetric cryptographic algorithms. You will also be made conversant with digital signatures and cryptographic hashing algorithms. In addition, you will also learn about the difference between hashing and encryption. Towards the end of this session, you will learn about using public keys as identities.

## Introduction to Cryptography

 [LO - Describe cryptography]



One of the most important components of blockchain is cryptography. Cryptography is a field that has been around for more than two thousand years. It is a technique that provides various methods that allow people or systems to keep things confidential. The major objective of cryptography is to ensure confidentiality, maintain integrity, and authenticate data. In simple words, it guarantees the confidentiality and integrity of the CIA (Confidentiality, Integrity, and Availability) triad; however, it does not guarantee data availability.



Any information in the form of a text, numeric data or computer program that can be read by humans or machine is known as plain text. When any plain text is encrypted using an algorithm and a key, it produces ciphertext. This process is known as encryption. The ciphertext is then transmitted to the intended recipient who decrypts the ciphertext using the key and the algorithm. This process is known as decryption. This process of encryption and decryption is known as cryptography. Figure 1 shows the working mechanism of cryptography:

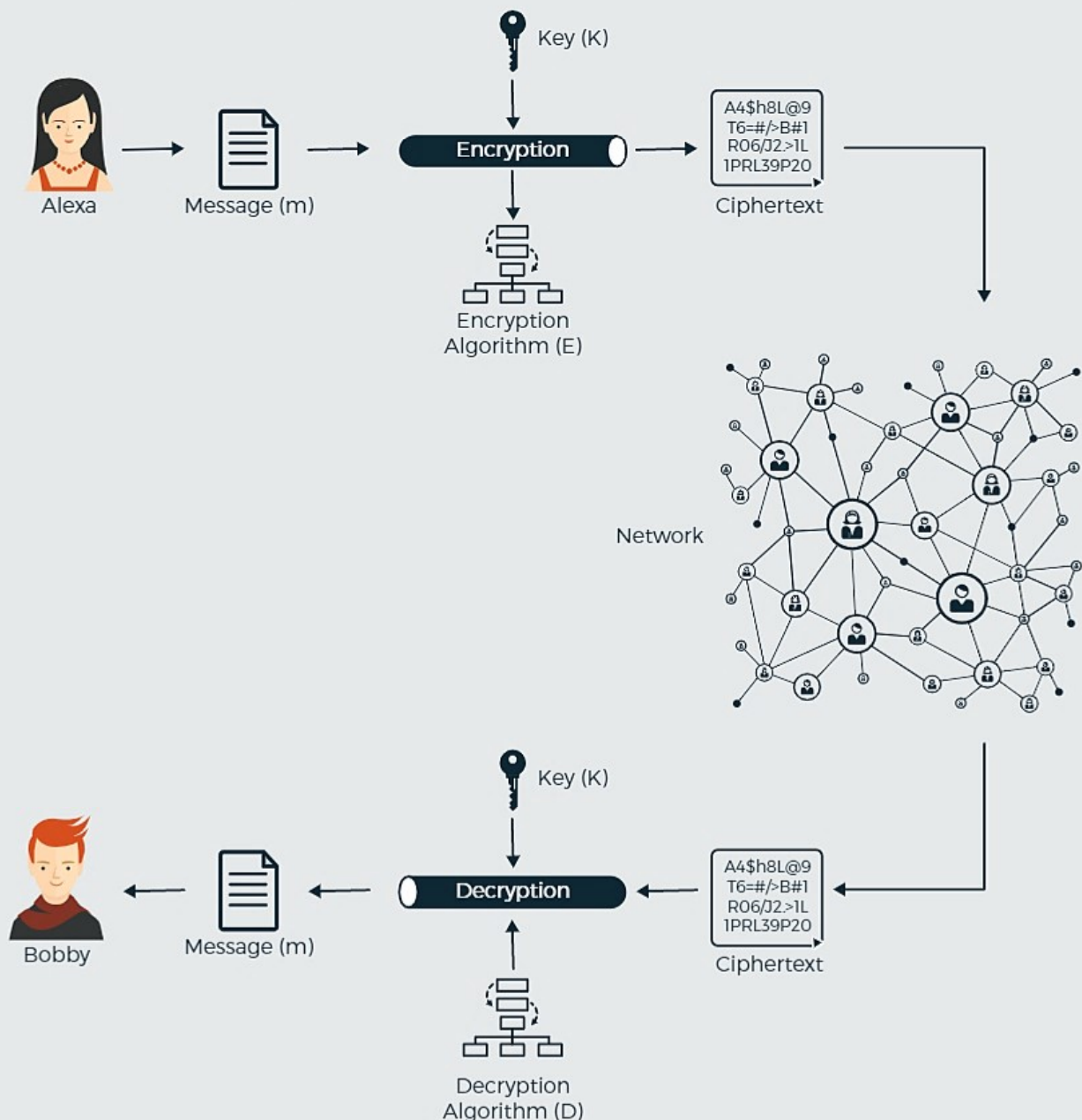


Figure 1: How Cryptography Works

Generally, there are two types of operations that are used to transform plain text into ciphertext— substitution and transposition.

A substitution cipher is defined as an encryption method that changes or swaps one character for another. It is an easy method of encryption. Caesar cipher is one of the oldest known substitution ciphers. It was used by Julius Caesar. In this cipher, the letters in the ciphertext are shifted to a certain number of spaces.

For example, for the following plain text:

### Encrypted using Caesar cipher

If every letter is shifted right by three letters, the ciphertext will be:

**Hqfubswhg xvlqj Fdhvdu flskhu**

Now, substitution ciphers can be easily cracked by analyzing the letter and word frequency. Another more complex variation of a substitution cipher is polyalphabetic substitution. It is a method in which every occurrence of an alphabet is replaced by another alphabet. For instance, Vigenère cipher is a cipher that uses a keyword to find the corresponding ciphertext in a table; this table is called Vigenère table.

A transposition cipher is a method that changes the order of characters of a word by using some predetermined method. Some of the examples of transposition ciphers are Columnar, Rail Fence and Route ciphers.

Cryptography includes various low-level algorithms which are used to create cryptographic protocols used by various applications. They are mainly the building blocks of a cryptographic system. Figure 2 shows these cryptography primitives:

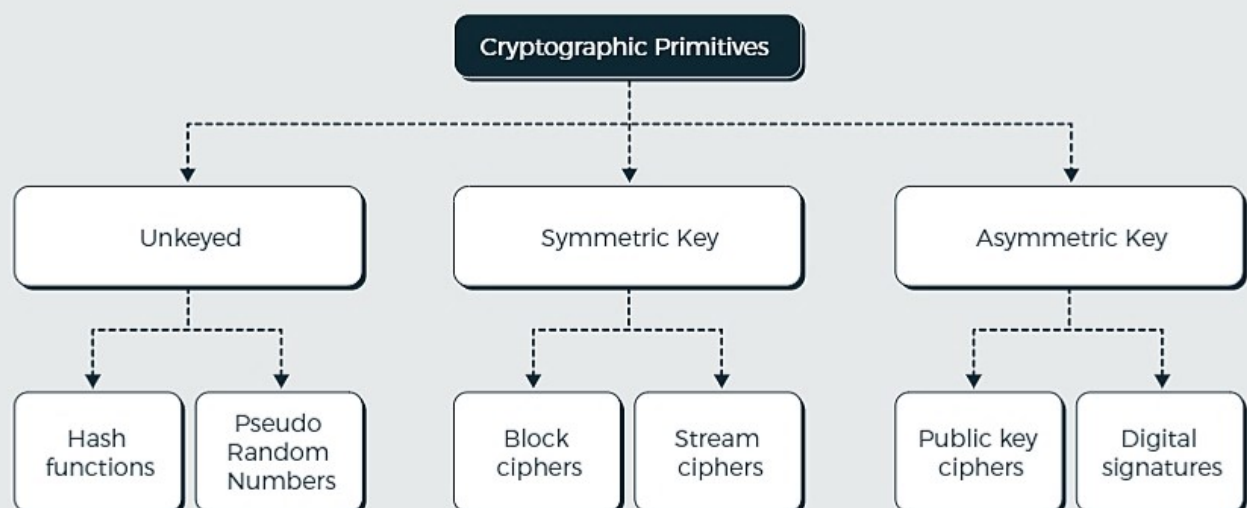


Figure 2: Cryptography Primitives



Now, let's discuss some of these primitives in the following sections.

## Exploring Symmetric Key Cryptography



[LO - Explain symmetric key cryptography]

Symmetric key cryptography is a key-based cryptography where the algorithms use the same key to perform encryption of plaintext and decryption of the ciphertext. The keys are shared between the two parties over a secure channel. Symmetric key cryptographic algorithms provide only confidentiality, and they do not provide authentication and non-repudiation. The origin of the message cannot be determined because of the use of the same key at both ends. Figure 3 shows the working of symmetric key cryptography:

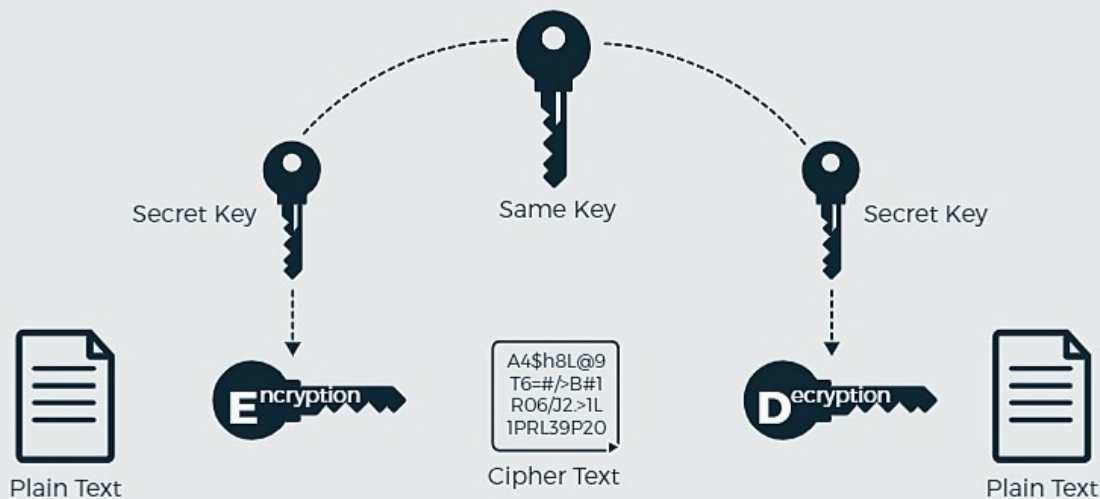


Figure 3: Symmetric Key Cryptography

Symmetric key cryptography does not have any kind of significant role in the blockchain. However, it helps in understanding the asymmetric cryptography.

A symmetric cipher can be encrypted either as stream encrypted or block encrypted. Let's understand them in the following sections.

### Stream Cipher

A stream cipher encrypts data bit by bit using keystream generators. The generators build a bit stream by using the XOR operation with the plain text bits. Since the stream cipher operates only one bit at a time, it is called One-Time Pad. For a secure implementation of a cryptographic system using the stream cipher, it is necessary that the generator should be unpredictable, and the key should not be reused. Rivest Cipher 4 (RC4) is one of the popular stream ciphers.

## Block Cipher

A block cipher encrypts a fixed-sized data that contains “n” number of bits. In simple words, it encrypts one block of data at one time. Some of the common sizes of blocks are 64 bits, 128 bits and 256 bits. A 128-bits block cipher needs 128-bits of plain text to encrypt it into 128-bits of ciphertext. If a block has a lesser number of bits than the allowed size, it will be padded.

## Exploring Asymmetric Key Cryptography



[LO - Explain asymmetric key cryptography]

Asymmetric key cryptography uses a pair of keys known as a public/private pair. The public key is created from the private key and can be freely distributed to the other users. This type of cryptography was developed due to the problems in symmetric key cryptography. In symmetric key cryptography, a shared key is used for both encryption and decryption which is tough to be shared between participants.

In asymmetric key cryptography, a public key is generated from a randomly generated private key. It is not possible to calculate the private key from the public key. Asymmetric key cryptography provides confidentiality, authentication, integrity, and non-repudiation. Figure 4 shows the working of asymmetric key cryptography:

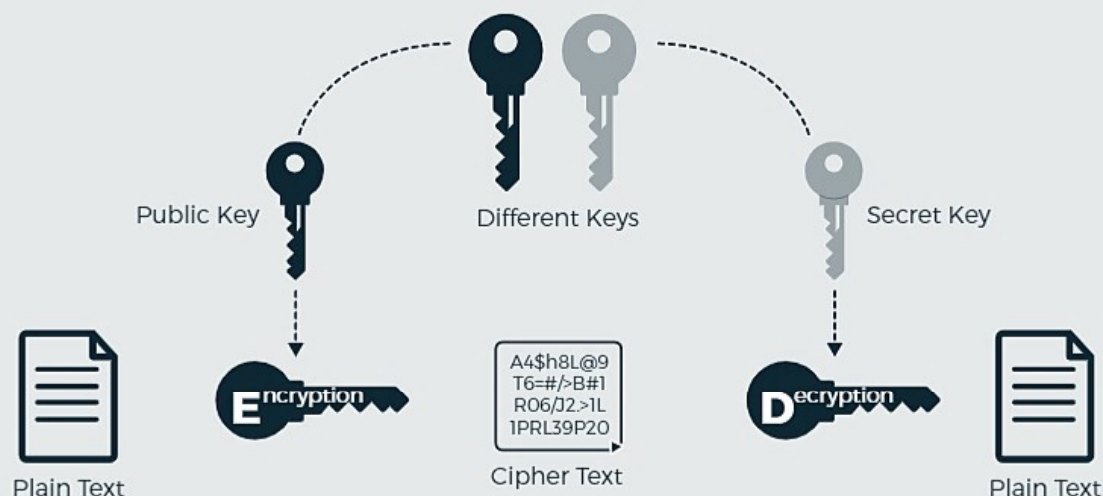


Figure 4: Asymmetric Key Cryptography

As shown in Figure 4, there is no secure channel required to share the keys. The algorithms used for encryption and decryption are similar.

Digital signature is one of the main applications of asymmetric key cryptography. Using a digital signature, a user can sign a message using his/her private key and anyone possessing the public key can verify the authenticity of the message. In blockchain applications such as cryptocurrencies, digital signatures are used to sign transactions with the private key to prove ownership.

Now, let's learn some of the approaches of asymmetric key cryptography.



## RSA Cryptosystem

One of the early implementations of asymmetric key cryptography is RSA (Rivest, Shamir and Adleman). This algorithm is named after the three men who developed it. It is widely used for encryption and digital signatures. It uses the principle of prime factorization to generate a public-private key pair.

In RSA, one key is used to perform the encryption and the other key is used to perform decryption. If a public key is used for encryption, its corresponding private key will be used for decryption. Public and private key pair is generated with the help of two large prime numbers. It is based on the fact that as long as large prime numbers are being used, it is impossible to compute public key from private key.

Currently, RSA uses keys having a typical length of 1024 or 2048 bits. With the increase in the length of the keys, the computational overhead of the RSA cryptography will increase.

## Elliptic-Curve Cryptography (ECC)

Elliptic-curve cryptography is an asymmetric key encryption technique which uses an elliptic curve equation to generate smaller and more secure cryptographic keys. An elliptic curve equation is as follows:

$$y^2 = x^3 + ax + b$$

where  $4a^3 + 27b^2 \neq 0$

Figure 5 shows an elliptic curve:

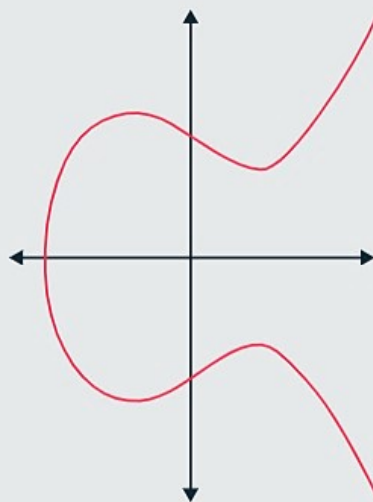


Figure 5: An Elliptic Curve

In ECC, the public-private key pair is generated on the basis of the properties of the elliptic curve equation instead of the conventional method of generating keys by multiplying prime numbers.

Suppose,  $G$  is an  $(x, y)$  point on a curve to which Alexa and Bobby has agreed. A random number ( $n$ ) will be the private key of Bobby and  $P$  will be the public key. The value of public key will be determined by multiplying  $n$  and  $G$  together:

$$P = n * G$$

The challenge with ECC is that if  $n$  is a 256-bit random number, it will be impossible to find its value even though  $G$  and  $P$  are known.


ECC requires keys of smaller size than RSA. According to experts, ECC can provide the same level of security with a 164-bit key for which other systems need a 1,024-bit key. It is mainly used in key exchange mechanisms and digital signatures. It is also used in Bitcoin's addressing system and in transaction signing operations. It is also used in some other blockchain applications.

## **Public Key Infrastructure (PKI)**

The problem of key distribution which existed in symmetric key cryptography does not exist in asymmetric key cryptography. However, there is one significant problem with asymmetric key cryptography is that there is no way to ensure that the public key used to encrypt the message is really the public key of the intended recipient and not of an intruder or eavesdropper. To solve this problem public key infrastructure (PKI) is introduced. With PKIs, it becomes possible to attest the user identity, thereby proving the authenticity of public keys. PKIs verify public keys by embedding them in a security certificate by digitally signing them.

A PKI system is used to register users, issue certificates, revoke, store and maintain those certificates. A public key certificate is in the format of X.509 and contains a set of attributes that are used to create a connection between the public key and the entity. The certificates of certificate authorities (CAs) are updated in the browsers of various operating systems automatically. However, due to centralized control, there are chances that certificates in the browsers may not get updated timely and this may result in higher security risks.

## **Exploring Cryptographic Hashing**

 [LO - Explain cryptographic hashing]



Hashing transforms data or a string of characters into a short and fixed length value which represents the original data. This value is called hash or message digest or hash key, which is unique.

Different hashing algorithms are used to perform hashing. These algorithms are also known as cryptographic hash functions. A hash function is a mathematical function that converts the input data into output, which is in encrypted form.



Figure 6 shows cryptographic hashing:



Figure 6: Cryptographic Hashing

The input data may vary in length, but the length of the output data is always fixed. This fixed length output data is known as a message digest. The fixed length of the message digest makes processing and storing quicker.

In everyday context, you encounter hashing when you create your email address. Your email provider, such as Gmail, does not save your password. Instead, it saves your password in the hash form after running your password through a hashing algorithm. When you log in and type your password, the hash of that input is compared to the stored hash. This way, without knowing your password, Google is able to verify your identity. Likewise, cryptographic hashing is used in the e-commerce sector and most importantly, in blockchain.

In blockchain, cryptographic hashing is used to conduct mining. It is also used to record new transactions with timestamp, and eventually to add a reference in the previous block. One-way nature of cryptographic hashing makes it essential for maintaining the integrity of a blockchain.

Hashing creates a unique identity string for every block by calculating its hash value. Each block maintains the hash value of the previous block and thus creates a form of a chain.

A cryptographic hash function is considered good if it has the following properties:

- **Pre-image resistance:** A pre-image resistance means that for a given hash value  $X$ , it should be infeasible to calculate message  $Y$  such that  $X = \text{hash}(Y)$ . This property refers to the one-way ness of a hash function.  
Pre-image resistance property enables a cryptographic hash function to hide any possible information/clue about the input.
- **Second pre-image resistance (weak collision resistance):** A weak collision resistance means that for a given message  $Y_1$ , it should be difficult to find another message  $Y_2$  such that  $\text{hash}(Y_1) = \text{hash}(Y_2)$ . It should be infeasible to find a second message having the same hash value as the first message.
- **Collision resistance:** A strong collision resistance means that it should be difficult to find two messages  $Y_1$  and  $Y_2$  in such a way that  $\text{hash}(Y_1) = \text{hash}(Y_2)$ . It should be infeasible to find two messages in such a way that their hash values are the same.



## ■ Hashing Algorithms

Hashing is an important cryptographic technique that ensures the integrity of data. To perform hashing, the sender calculates a hash value on the basis of the data. In hashing, two different messages cannot have the same hash value. When the receiver receives the hash value, it runs the message through the same hashing algorithm that has been used by the sender. If the created hash value is similar to the received hash value, it proves that the integrity of the message is intact.

Hashing algorithms are categorized on the basis of their implementations, digest size and other things. Some of the algorithms are:


- **Message Digest (MD):** MD function is a family of hash functions developed by Ron Rivest. The MD function is a 128-bit hash function. Its family consists of MD2, MD4, MD5, MD6. MD5 is a widely used hash function that results in 128-bit hash value. MD5 function is vulnerable to collision attacks. Therefore, it is not recommended by professionals.
- **Secure Hashing Algorithms (SHA):** The US National Institute of Standards and Technology (NIST) published four SHA hash functions. They are SHA-0, SHA-1, SHA-2, and SHA-3. SHA-0 is the first function of the SHA family. The SHA-0 and SHA-1 are 160-bit hash functions that perform 80 rounds of computations. The SHA-1 function was developed in 1993 to overcome the drawbacks of SHA-0. The SHA-1 function is mostly used in protocols such as SSL and applications.  
  
The SHA-2 is one of the strong hash functions. It contains four functions that include SHA-224, SHA-256, SHA-384, and SHA-512. SHA-224, SHA-256, SHA-384, and SHA-512 generate 224, 256, 384, and 512 bits after performing 64, 64, 80, and 80 rounds of computations, respectively. The SHA-224 and SHA-256 functions perform calculations on 512-bit blocks whereas the SHA-384 and SHA-512 functions perform calculations on 1024-bit blocks. The SHA-3 function, which was developed in 2014, is also a family of hash functions. It generates hash values ranging from 224 to 512 bits after performing 120 rounds of computations.
- **Keccak 256/512:** Keccak is a family of cryptographic hash algorithms that won the SHA-3 competition organized by NIST. It is used for authentication, encryption and pseudo-random number generation. It uses sponge construction and Keccak-f cryptographic permutation. Sponge construction is based on a random function or random permutation that allows inputting and outputting any amount of data, thereby providing better flexibility.

After Keccak won the competition, NIST adjusted some of the parameters of Keccak to improve its efficiency. However, the Ethereum Foundation decided to implement the original Keccak algorithm, as proposed by its inventors, rather than implementing the SHA-3 standard as modified by NIST.





## | Exploring Digital Signatures

 [LO - Describe digital signatures]

A digital signature is a widely used cryptography technique which provides a high level of security. A digital signature validates the authenticity of any digital document, software, or message by using a hash value which is encrypted through the private key of the sender. It associates a digital signer with a document or message. It provides authentication, integrity, and nonrepudiation of communication across the Internet.

When digital signatures are used in conventional banking, a CA is used as a trusted service provider that ensures key security and provides digital certificates containing digital signatures. However, since Bitcoin is based on decentralization, the communication is only authenticated by the content of the transaction itself. When a user requests for the payment of a transaction, he/she includes a signature of the payer of the transaction. Since no one can create a signature without the private key, this proves the validity of the message.

When you request a payment, it includes a signature that could only have been created by the owner of the address that held the money previously. Nobody can change the message, because the signature would be invalid for any other message, and they cannot produce another valid signature without having access to the private key. The secret is not revealed to order the payment. Therefore, it is not necessary to encrypt communication within the network, and also there is no need to certify the identity of network participants. One can tell directly from the message whether it is valid or not, making the identity of the source of the message irrelevant.

Digital signature is based on asymmetric key cryptography or public key cryptography that uses two keys, public and private. For creating a digital signature, the signer or sender needs to find a hash value for the required message or data and encrypt it using his or her private key.

Now, the sender attaches this hash value and a copy of his or her public key within a certificate and sends it to the receiver. After receiving the signature, the receiver verifies it by separating the encrypted hash, message, and the certificate. The receiver also finds the hash value of the message and decrypts it using the public key. After this, the receiver compares both the hash values. If the values are the same, then the message has not changed; however, if they are different, it would mean that the message has changed. Figure 7 illustrates the signing and verification process of a digital signature:

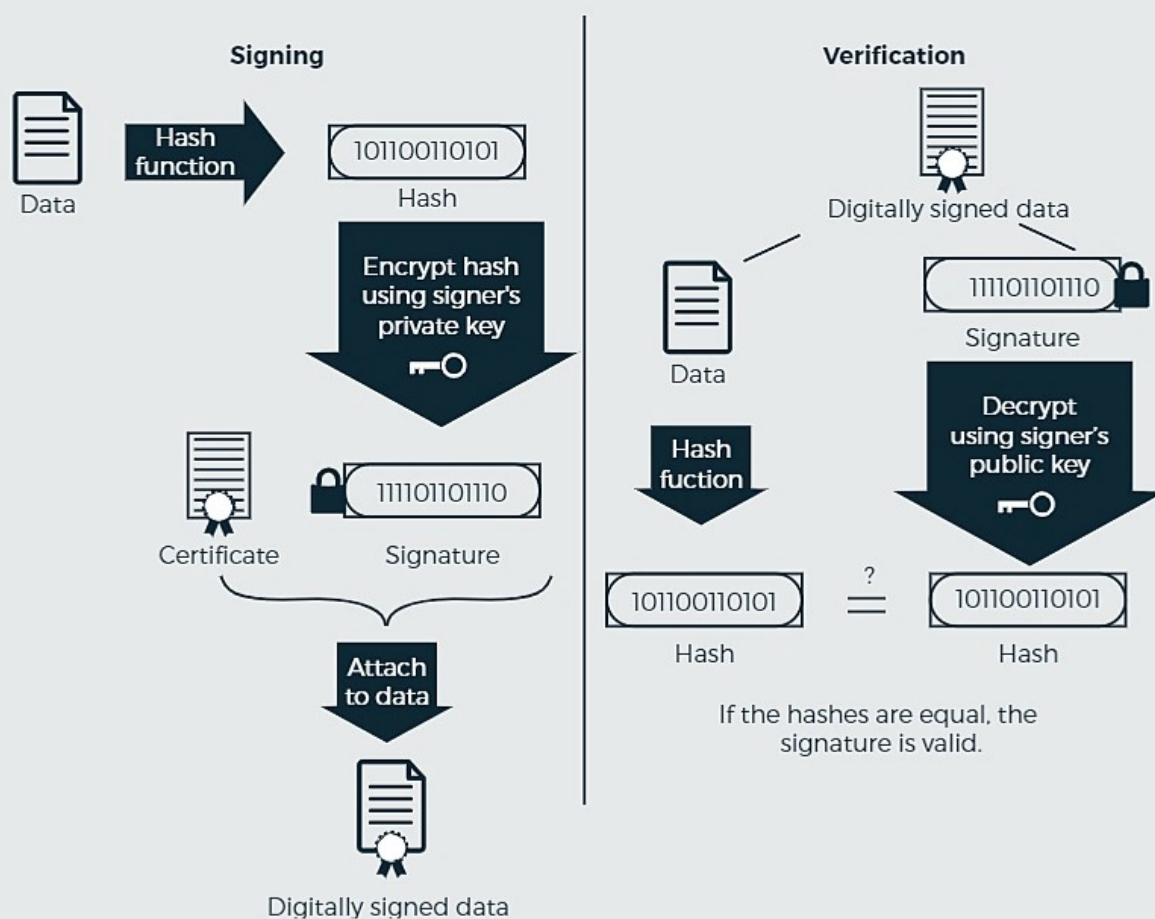


Figure 7: Signing and Verification of Digital Signatures

## Difference Between Hashing and Encryption

 [LO - Explain the difference between hashing and encryption]




In cryptographic hashing, a mathematical function is used that takes data as input and then converts that data into a fixed length irreversible string known as message digest. Hashing is a one-way process while encryption is a two-way or reversible process.

Another difference between hashing and encryption is that hashing will always return a fixed length output regardless of the input size. However, in encryption, the size of the output depends on the size of input.





## | Using Public Keys as Identities

 [LO - Discuss how to use public keys as identities]

Using public keys as an identity means that the public key can be used to identify an entity. It becomes possible by using digital signatures. Using digital signatures, one can verify the validity of a message based on the public key, message, and signature.

In a real-world scenario, your bank account number or your social security number is considered as a person's identity. Similarly, in blockchain, a signed message uploaded by an entity can only be verified with the entity's public key. Therefore, when a signature is verified correctly with someone's public key, it can be implied that the public key is the actor or identity that makes statements valid by signing them. Therefore, in case, if a public key is not able to verify a signed message, then that message is considered as invalid.

A message can be signed using the private key and then can be published to the network. However, the messages are being published on that network on behalf of that public key.

Now, the disadvantage of using public keys as identities is that anyone can make a new identity whenever they want. They are only required to register on a central management interface which generates a random public or private key pair. The solution to this problem is provided by decentralized identity management.

Now, instead of registering at a central interface, it is not necessary to have a username. Anyone can make a new identity at any time.

# SUMMARY

- The process of encryption and decryption is known as cryptography.
- There are two types of operations that are used to transform plain text into ciphertext— substitution and transposition.
- In symmetric key cryptography, the same key is used to perform encryption of plaintext and decryption of the ciphertext.
- In asymmetric key cryptography, a pair of keys known as a public/private pair is used to perform encryption of plaintext and decryption of the ciphertext.
- A digital signature validates the authenticity of any digital document, software, or message by using a hash value.
- A hash function is a mathematical function that converts the input data into encrypted data as output.
- When a signature is verified correctly with someone's public key, then it can be implied that the public key is the actor that makes statements valid by signing them.