UNIT-I

Concepts

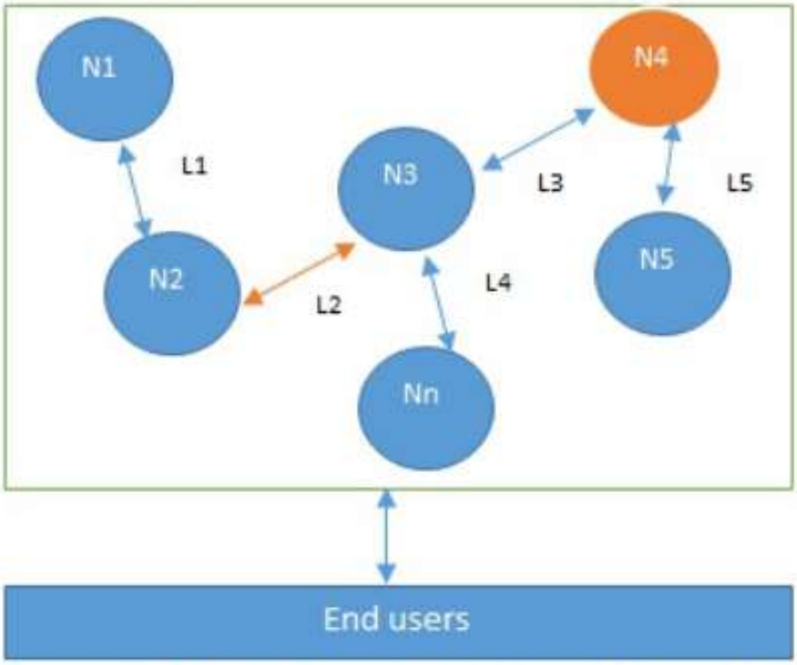
Blockchain 101: Distributed systems, History of Blockchain and bitcoin, Introduction to Blockchain: Blockchain architecture, Generic elements of a blockchain, How blockchain works, Merkle Trees, Tiers and Types of Blockchain, Features, Benefits and Limitations of Blockchain , Consensus Protocols, Types of consensus mechanisms

Distributed Systems

Distributed systems are a computing paradigm whereby two or more nodes work with each other in a coordinated fashion to achieve a common outcome. It is modelled in such a way that end users see it as a single logical platform. For example, Google's search engine is based on a large distributed system, but to a user, it looks like a single, coherent platform.

A node can be defined as an individual player in a distributed system. All nodes are capable of sending and receiving messages to and from each other. Nodes can be honest, faulty, or malicious, and they have memory and a processor. A node that exhibits irrational behavior is also known as a Byzantine node after the Byzantine Generals Problem.

A small-scale example of a distributed system is shown in the following diagram. This distributed system has six nodes out of which one (N4) is a Byzantine node leading to possible data inconsistency. L2 is a link that is broken or slow, and this can lead to partition in the network.



Design of a distributed system: N4 is a Byzantine node, L2 is broken or a slow network link

The primary challenge in distributed system design is coordination between nodes and fault tolerance. Even if some of the nodes become faulty or network links break, the distributed system should be able to tolerate this and continue to work to achieve the desired result.

History of Blockchain and Bitcoin

Blockchain was introduced with the invention of Bitcoin in 2008. Its practical implementation then occurred in 2009

Electronic cash: The concept of electronic cash or digital currency is not new. Since the 1980s, e-cash protocols have existed that are based on a model proposed by David Chaum.

Two fundamental e-cash system issues need to be addressed: accountability and anonymity.

Accountability is required to ensure that cash is spendable only once (double-spend problem) and that it can only be spent by its rightful owner. Double spend problem arises when same money can be spent twice. As it is quite easy to make copies of digital data, this becomes a big issue in digital currencies as you can make many copies of same digital cash. Anonymity is required to protect users' privacy.

In 2009, the first practical implementation of an electronic cash (e-cash) system named Bitcoin appeared. The term cryptocurrency emerged later. For the very first time, it solved the problem of distributed consensus in a trustless network. It used public key cryptography with a Proof of Work (PoW) mechanism to provide a secure, controlled, and decentralized method of minting digital currency.

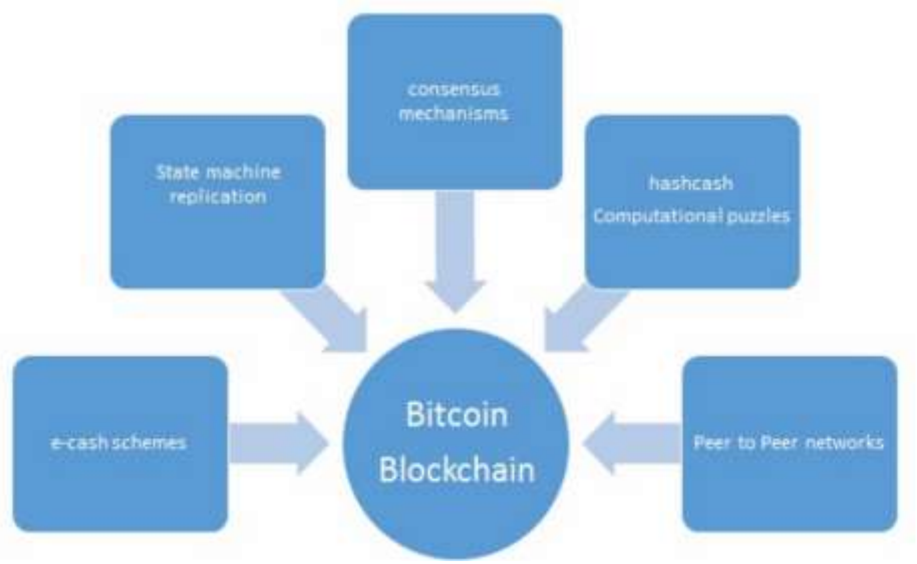


Figure :The various ideas that supported the invention of Bitcoin and blockchain

<u>Introduction to blockchain</u>

In 2008, a groundbreaking paper entitled Bitcoin: A Peer-to-Peer Electronic Cash System was written on the topic of peer-to-peer electronic cash under the pseudonym Satoshi Nakamoto. It introduced the term chain of blocks.

Definition

Layman's definition: Blockchain is an ever-growing, secure, shared record keeping system in which each user of the data holds a copy of the records, which can only be updated if all parties involved in a transaction agree to update.

Technical definition: Blockchain is a peer-to-peer, distributed ledger that is cryptographically-secure, append-only, immutable (extremely hard to change), and updateable only via consensus or agreement among peers.

Features of bloclchain

Peer-to-peer : This means that there is no central controller in the network, and all participants talk to each other directly. This property allows for cash transactions to be exchanged directly among the peers without a third-party involvement, such as by a bank.

Distributed ledger Blockchain is a distributed ledger, which simply means that a ledger is spread across the network among all peers in the network, and each peer holds a copy of the complete ledger.

Cryptographically-secure Means cryptography has been used to provide security services which make this ledger secure against tampering and misuse. These services include non-repudiation, data integrity, and data origin authentication.

Append-only (Immutable): which means that data can only be added to the blockchain in time-ordered sequential order. This property implies that once data is added to the blockchain, it is almost impossible to change that data and can be considered practically immutable.

Uniqueness: This blockchain feature ensures that every transaction is unique and has not already been spent (double-spend problem). This feature is especially relevant with cryptocurrencies, where detection and avoidance of double spending are a vital requirement.

Updateable via consensus

- Finally, the most critical attribute of a blockchain is that it is updateable only via consensus. This is what gives it the power of decentralization.
- In this scenario, no central authority is in control of updating the ledger.
- Instead, any update made to the blockchain is validated against strict criteria defined by the blockchain protocol and added to the blockchain only after a consensus has been reached among all participating peers/nodes on the network.
- To achieve consensus, there are various consensus facilitation algorithms which ensure that all parties are in agreement about the final state of the data on the blockchain network and resolutely agree upon it to be true.

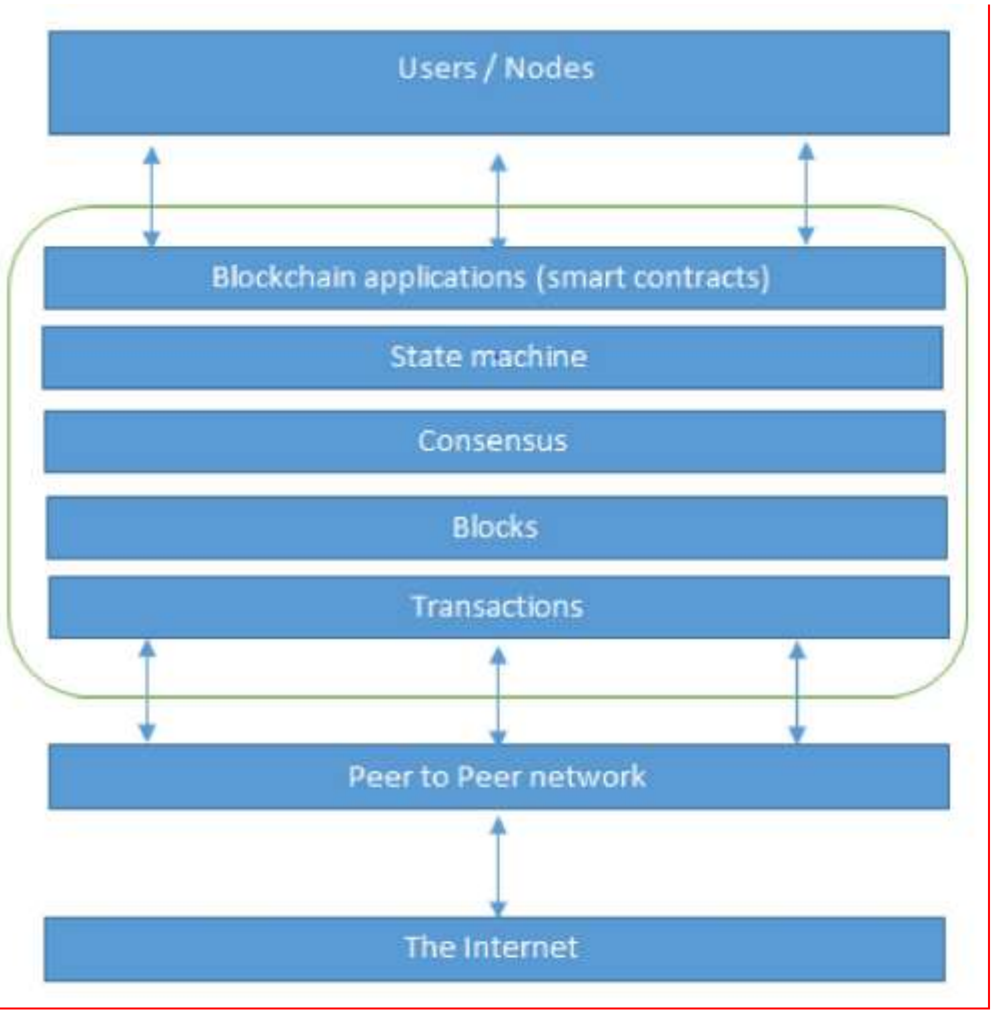


Figure: The network view of a blockchain

Generic structure of a block

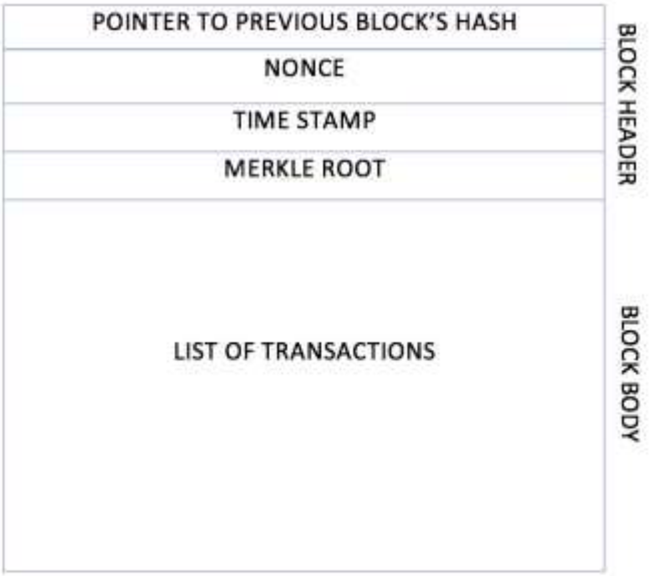
A **block** is merely a selection of transactions bundled together and organized logically. A block is made up of transactions, and its size varies depending on the type and design of the blockchain in use.

A **transaction** is a record of an event, for example, the event of transferring cash from a sender's account to a beneficiary's account.

A **genesis block** is the first block in the blockchain that is hardcoded at the time the blockchain was first started.

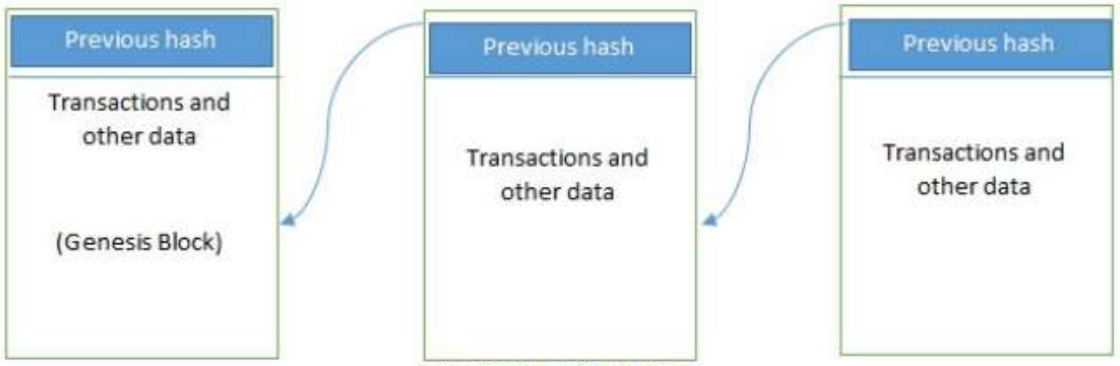
A **nonce** is a number that is generated and used only once. A nonce is used extensively in many cryptographic operations to provide replay protection, authentication, and encryption.

Merkle root is a hash of all of the nodes of a Merkle tree. Merkle trees are widely used to validate the large data structures securely and efficiently. In the blockchain world, Merkle trees are commonly used to allow efficient verification of transactions. Merkle root in a blockchain is present in the block header section of a block, which is the hash of all transactions in a block. This means that verifying only the Merkle root is required to verify all transactions present in the Merkle tree instead of verifying all transactions one by one.



The generic structure of a block.

Generic elements of a blockchain


Previous hash
Transactions and other data
(Genesis Block)
Previous hash
Transactions and other data
Previous hash
Transactions and other data

Generic structure of a blockchain

Address: Addresses are unique identifiers used in a blockchain transaction to denote senders and recipients. An address is usually a public key or derived from a public key. While addresses can be reused by the same user, addresses themselves are unique.

Transaction: A transaction is the fundamental unit of a blockchain. A transaction represents a transfer of value from one address to another.

Block: A block is composed of multiple transactions and other elements, such as the previous block hash (hash pointer), timestamp, and nonce.

Peer-to-peer network: As the name implies, a peer-to-peer network is a network topology wherein all peers can communicate with each other and send and receive messages.

State machine: A blockchain can be viewed as a state transition mechanism whereby a state is modified from its initial form to the next one and eventually to a final form by nodes on the blockchain network as a result of a transaction execution, validation, and finalization process.

Node: A node in a blockchain network performs various functions depending on the role that it takes on. A node can propose and validate transactions and perform mining to facilitate consensus and secure the blockchain.

How blockchain works

1. A node starts a transaction by first creating and then digitally signing it with its private key. Most commonly this is a data structure that represents transfer of value between users on the blockchain network.
2. 2. A transaction is propagated (flooded) by using a flooding protocol, called Gossip protocol, to peers that validate the transaction based on preset criteria. Usually, more than one node are required to verify the transaction.
3. Once the transaction is validated, it is included in a block, which is then propagated onto the network. At this point, the transaction is considered confirmed.
4. The newly-created block now becomes part of the ledger, and the next block links itself cryptographically back to this block. This link is a hash pointer. At this stage, the transaction gets its second confirmation and the block gets its first confirmation.
5. Transactions are then reconfirmed every time a new block is created.

Tiers of Blockchain:

Blockchain 1.0: This tier was introduced with the invention of Bitcoin, and it is primarily used for cryptocurrencies. Also, as Bitcoin was the first implementation of cryptocurrencies, it makes sense to categorize this first generation of blockchain technology to include only cryptographic currencies. This generation started in 2009 when Bitcoin was released and ended in early 2010.

Blockchain 2.0: This second blockchain generation is used by financial services and smart contracts. This tier includes various financial assets, such as derivatives, options, swaps, and bonds. Applications that go beyond currency, finance, and markets are incorporated at this tier. Ethereum, Hyperledger, and other newer blockchain platforms are considered part of Blockchain 2.0. This generation started when ideas related to using blockchain for other purposes started to emerge in 2010.

Blockchain 3.0: This third blockchain generation is used to implement applications beyond the financial services industry and is used in government, health, media, the arts, and justice. This generation of blockchain emerged around 2012 when multiple applications of blockchain technology in different industries were researched.

Blockchain X.0: This generation represents a vision of blockchain singularity where one day there will be a public blockchain service available that anyone can use just like the Google search engine. It will provide services for all realms of society. It will be a public and open distributed ledger with general-purpose rational agents (Machina economicus) running on a blockchain, making decisions, and interacting with other intelligent autonomous agents on behalf of people, and regulated by code instead of law or paper contracts.

Types of Blockchain Technology

Public blockchains: public blockchains are not owned by anyone. They are open to the public, and anyone can participate as a node in the decision-making process. Users may or may not be rewarded for their participation. All users of these permissionless or unpermissioned ledgers maintain a copy of the ledger on their local nodes and use a distributed consensus mechanism to decide the eventual state of the ledger. Bitcoin and Ethereum are both considered public blockchains

Private blockchains As the name implies, private blockchains are just that—private. That is, they are open only to a consortium or group of individuals or organizations who have decided to share the ledger among themselves. There are various blockchains now available in this category, such as HydraChain and Quorum

Consortium blockchain is midway between public and private blockchain, as it involves the blockchain of not just an individual organization, but a cluster of them belonging to different organizations. Yet it still exists still within a controlled group of users, unlike a public distributed ledger.

Benefits and limitations of blockchain

Benefits

Decentralization: This is a core concept and benefit of the blockchain. There is no need for a trusted third party or intermediary to validate transactions; instead, a consensus mechanism is used to agree on the validity of transactions.

Transparency and trust: Because blockchains are shared and everyone can see what is on the blockchain, this allows the system to be transparent. As a result, trust is established.

Immutability: Once the data has been written to the blockchain, it is extremely difficult to change it back. It is not genuinely immutable, but because changing data is so challenging and nearly impossible, this is seen as a benefit to maintaining an immutable ledger of transactions.

High availability: As the system is based on thousands of nodes in a peer-to-peer network, and the data is replicated and updated on every node, the system becomes highly available. Even if some nodes leave the network or become inaccessible, the network as a whole continues to work, thus making it highly available.

Highly secure: All transactions on a blockchain are cryptographically secured and thus provide network integrity.

Simplification of current paradigms: The current blockchain model in many industries, such as finance or health, is somewhat disorganized. In this model, multiple entities maintain their own databases and data sharing can become very difficult due to the disparate nature of the systems. However, as a blockchain can serve as a single shared ledger among many interested parties, this can result in simplifying the model by reducing the complexity of managing the separate systems maintained by each entity.

. **Cost saving:** As no trusted third party or clearing house is required in the blockchain model, this can massively eliminate overhead costs in the form of the fees which are paid to such parties.

Problems

Scalability: Blockchain like bitcoin has consensus mechanisms which require every participating node to verify the transaction. It limits the number of transactions a blockchain network can process. So bitcoin was not developed to do the large scale volumes of transactions that many of the other institutions are doing. Currently, bitcoin can process a maximum of **seven transactions per second**.

Adaptability

Lack of Awareness

There is a lot of discussion about blockchain, but people do not know the true value of blockchain and how they could implement it in different situations.

Privacy

Consensus Mechanisms

Consensus is a process of agreement between distrusting nodes on the final state of data. To achieve consensus, different algorithms are used. It is easy to reach an agreement between two nodes (in client-server systems, for example), but when multiple nodes are participating in a distributed system and they need to agree on a single value, it becomes quite a challenge to achieve consensus. This process of attaining agreement common state or value among multiple nodes despite the failure of some nodes is known as **distributed consensus**.

There are various requirements that must be met to provide the desired results in a consensus mechanism. The following describes these requirements:
Agreement: All honest nodes decide on the same value
Termination: All honest nodes terminate execution of the consensus process and eventually reach a decision

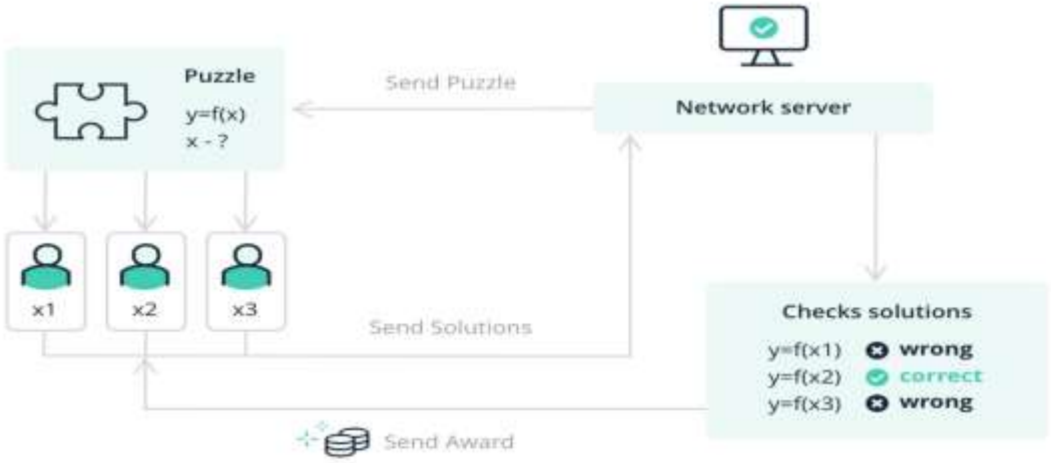
Validity: The value agreed upon by all honest nodes must be the same as the initial value proposed by at least one honest node

Fault tolerant: The consensus algorithm should be able to run in the presence of faulty or malicious nodes (Byzantine nodes)

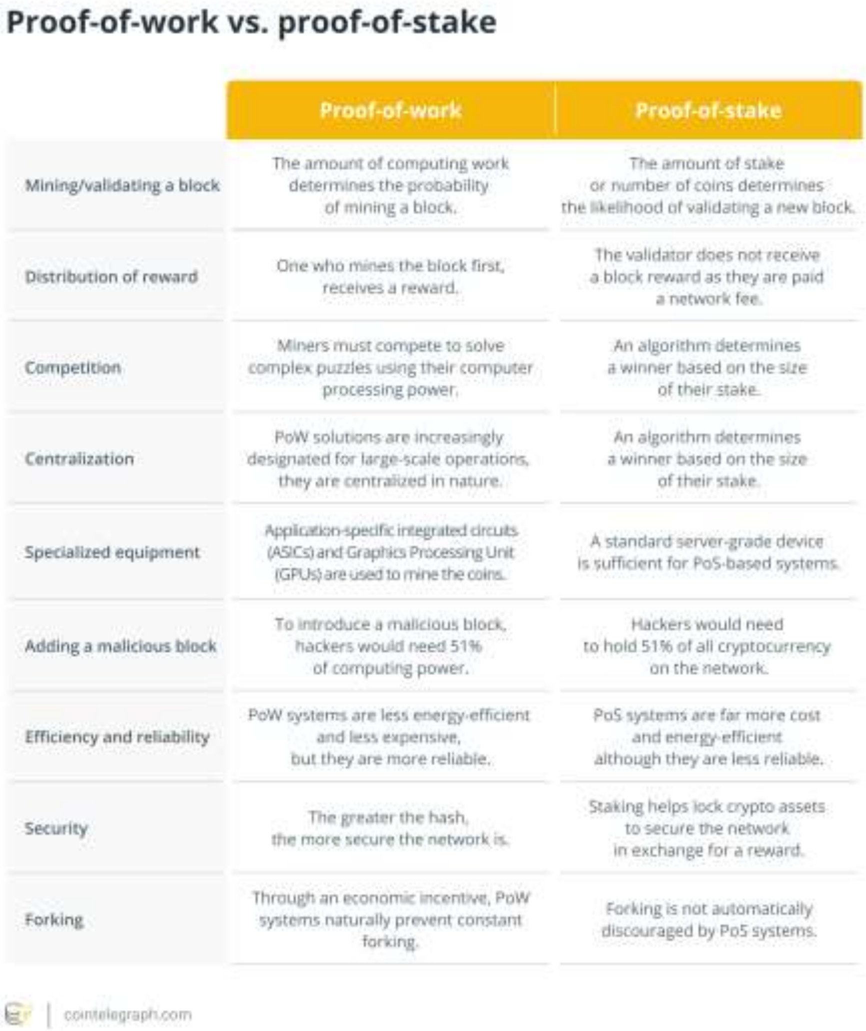
Integrity: This is a requirement that no node can make the decision more than once in a single consensus cycle

<u>Types of Consensus Mechanisms</u>

PoW(Proof : A consensus mechanism in which computing power is used to verify cryptocurrency transactions and add them to the blockchain.



PoS(Proof of Stake): The ability to mine is determined by how many tokens of this currency the user owns. In PoS miner does not earn rewards but is paid with network fees.



Proof-of-work vs. proof-of-stake

	Proof-of-work	Proof-of-stake
Mining/validating a block	The amount of computing work determines the probability of mining a block.	The amount of stake or number of coins determines the likelihood of validating a new block.
Distribution of reward	One who mines the block first, receives a reward.	The validator does not receive a block reward as they are paid a network fee.
Competition	Miners must compete to solve complex puzzles using their computer processing power.	An algorithm determines a winner based on the size of their stake.
Centralization	PoW solutions are increasingly designated for large-scale operations, they are centralized in nature.	An algorithm determines a winner based on the size of their stake.
Specialized equipment	Application-specific integrated circuits (ASICs) and Graphics Processing Unit (GPUs) are used to mine the coins.	A standard server-grade device is sufficient for PoS-based systems.
Adding a malicious block	To introduce a malicious block, hackers would need 51% of computing power.	Hackers would need to hold 51% of all cryptocurrency on the network.
Efficiency and reliability	PoW systems are less energy-efficient and less expensive, but they are more reliable.	PoS systems are far more cost and energy-efficient although they are less reliable.
Security	The greater the hash, the more secure the network is.	Staking helps lock crypto assets to secure the network in exchange for a reward.
Forking	Through an economic incentive, PoW systems naturally prevent constant forking.	Forking is not automatically discouraged by PoS systems.

cointelegraph.com

Delegated Proof of Stake (DPoS): DPoS is a system in which a fixed number of elected entities (called *block producers* or *witnesses*) are selected to create blocks in a round-robin order. Block producers are voted into power by the users of the network, who each get a number of votes proportional to the number of tokens they own on the network (their *stake*).

Proof of elapsed time (PoET) is a blockchain network consensus technique that uses a fair lottery system to maintain process efficiency while preventing excessive resource and energy consumption.

Proof of Deposit (PoD): In this case, nodes that wish to participate in the network have to make a security deposit before they can mine and propose blocks. This mechanism is used in the Tendermint blockchain.

Proof of Importance : Proof of Importance is the mechanism that is used to determine which nodes in the network are eligible to add a block to the blockchain, by a process that is known as ‘harvesting’ or ‘vesting’ by NEM which stands for New Economy Movement which is a blockchain. In exchange for harvesting a block, nodes are able to collect the transaction fees within that block which the validator gets as a reward.

Reputation-based mechanisms: As the name suggests, a leader is elected by the reputation it has built over time on the network. It is based on the votes of other members.

PBFT(Practical Byzantine Fault Tolerant) : pBFT-enabled distributed system are sequentially ordered, with one node being the primary (or the leader node) and others referred to as secondary (or the backup nodes). Note here that any eligible node in the system can become the primary by transitioning from secondary to primary (typically, in the case of a primary node failure). The goal is that all honest nodes help in reaching a consensus regarding the state of the system using the majority rule.

Proof of Activity: Proof of Activity is a hybrid consensus mechanism that is a combination of two other Blockchain consensus mechanisms: Proof of Work (POW) and Proof of Stake (POS). It attempts to leverage the best of both PoW and PoS consensus mechanisms to validate and generate new blocks in the Blockchain.

Proof of Capacity (PoC): This scheme uses hard disk space as a resource to mine the blocks. This is different from PoW, where CPU resources are used. In in PoC, hard disk space is utilized for mining and as such is also known as hard drive mining.