P P SAVANI
UNIVERSITY

Module 6

Blockchain architecture

Outline

P P SAVANI UNIVERSITY

- Administrator (operator) Considerations
- Security: Public vs. Private Blockchains
- Architect Considerations
- Network Consensus Considerations

Peer to Peer System - Napster

P P SAVANI UNIVERSITY

- The music industry has worked for a long time in the following way: musicians made contracts with studios, which recorded the songs, produced and marketed the music records on a variety of media (e.g., vinyl, tape, or CD), which in turn were sold to the customers via a variety of distribution channels, including department stores and specialized shops.

- The studios actually worked as intermediaries between musicians and people who enjoy listening to music. Music studios could maintain their role as intermediaries due to their exclusive knowledge and skills in producing, marketing, and distributing records. However, in the first decade of the 2000s, the environment in which the music studios operated changed dramatically.

Peer to Peer System - Napster

P P SAVANI UNIVERSITY

- The digitalization of music, the availability of recording equipment at affordable prices, the growing spread of privately used PCs, and the emergence of the Internet made music studios dispensable. The three functions of music studios—**producing, marketing, and distributing** records—could be done by the artists and the consumers themselves. Napster played a major role in the replacement of the music studios as intermediaries. With Napster, people no longer relied on the music studios to get the latest hits. It was possible to share individual music files with people all over the world without the need to buy any CDs.
- The peer-to-peer approach of Napster, actually being a kind of a digital sharing bazaar for mp3 files, gave consumers access to a wider range of music than ever before, making the music studios partly dispensable and causing them significant losses.

The Potential of Peer-to-Peer Systems

P P SAVANI UNIVERSITY

- The power of peer-to-peer systems is not restricted to the music industry. Each industry that mainly acts as a middleman between producers and customers of immaterial or digital goods and services is vulnerable to being replaced by a peer-to-peer system. This statement may sound a bit abstract, but you may discover many middlemen for immaterial and digital goods and services around you once you recognize the largest of them all: the financial industry.

Centralized Peer to Peer

P P SAVANI UNIVERSITY

- construction since they are made of individual nodes that share their computational resources among others. However, there are also peer-to-peer systems that still utilize elements of centralization. Centralized peer-to-peer systems maintain central nodes to facilitate the interaction between peers, to maintain directories that describe the services offered by the peer nodes, or to perform look-ups and identification of the nodes.
- An example of a centralized peer-to-peer system is Napster, which maintained a central database of all nodes connected with the system and the songs available on these nodes.
- Systems are distributed computer systems by Peer-to-peer.

What exactly is a peer to peer system?

P P SAVANI UNIVERSITY

- Peer-to-peer systems are distributed software systems that consist of nodes (individual computers), which make their computational resources (e.g., processing power, storage capacity, or information distribution) directly available to another.
- When joining a peer-to-peer system, users turn their computers into nodes of the system that are equal concerning their rights and roles. Although users may differ with respect to the resources they contribute, all the nodes in the system have the same functional capability and responsibility.
- Hence, the computers of all users are both suppliers and consumers of resources.
- For example, in a peer-to-peer file sharing system, the individual files are stored on the users' machines. When someone wants to download a file in such a system, he or she is downloading it from another person's machine, which could be the next door neighbor or someone located halfway around the world.

The Link Between Peer-to-Peer Systems and the Blockchain

- The blockchain can be considered a tool for achieving and maintaining integrity in distributed systems. Purely distributed peer-to-peer systems may use the blockchain in order to achieve and to maintain system integrity.

- Hence, the link between purely distributed peer-to-peer systems and the blockchain is its usage for achieving and maintaining integrity in purely distributed systems.

The Link Between Peer-to-Peer Systems and the Blockchain

P P SAVANI UNIVERSITY

- The relation between purely distributed peer-to-peer systems to the blockchain is that the former uses the latter as a tool to achieve and maintain integrity.
- Hence, the argument that explains the excitement about and the potential of the blockchain is: Purely distributed peer-to- peer systems have a huge commercial potential as they can replace centralized systems and change whole industries due to disintermediation.
- Since purely distributed peer-to-peer systems may use the blockchain for achieving and maintaining integrity, the blockchain becomes important as well.
- However, the major fact that excites people is the disintermediation. The blockchain is only a means to an end that helps to achieve that.

Vulnerabilities of Peer to peer system

P P SAVANI UNIVERSITY

- Counterfeiting bank notes is a severe crime in any country because it undermines the foundation and functioning of the economy by creating purchasing power that is not backed up by valuable resources.
- As a result, most bank notes are equipped with security features that make counterfeiting impossible or prohibitively costly at least.
- These security features, such as unique numbers, watermarks, or fluorescent fibers, work well with physical bank notes and other physical goods. But what happens if money or goods become digital and are managed in distributed peer-to-peer systems of ledgers? This step explains a specific vulnerability of distributed peer-to-peer systems used for managing ownership that is equivalent to counterfeiting bank notes. As it turns out, this vulnerability is a prominent example of violated system integrity.

Vulnerabilities of Peer to peer system

P P SAVANI UNIVERSITY

- Let's consider a peer-to-peer system for managing ownership of real estate. In such a system, the ledgers that keep track of ownership information are maintained by the individual computers of its members instead of being maintained in a central database.
- Hence, each peer maintains his or her own copy of the ledger. As soon as the ownership of a house is transferred from one person to another, all the ledgers of the system need to be updated in order to contain the latest version of reality. However, passing information forward among peers and updating the individual ledgers require time.
- Until the last member of the system receives the new information and updates his or her copy of the ledger, the system will not be consistent. Some peers already know about the latest transfer of ownership, while other peers have not yet received that information. The fact that not all ledgers have up-to-date information makes them prone to be exploited by anyone who already has the

The Double Spending Problem

P P SAVANI UNIVERSITY

- Let's also imagine the following situation. Person A sells his house to person B. The transfer of ownership from A to B is documented in one of the ledgers in the peer-to-peer system.
- This particular ledger needs to inform other peers about this transfer, who in turn inform other peers as well, until eventually all peers learn about the transfer of ownership from A to B.
- However, suppose that person A quickly approaches another ledger of the system and demands to document a different transfer of ownership of the identical house: the sale from person A to person C.
- If this peer has not yet learned about the transfer of ownership from A to B that happened in the past, this peer will approve and document the transfer of ownership from A to C for the identical house.

Continue…

- Hence, A was able to sell his house twice by exploiting the fact that distributing information about his first sell requires time.

- But B and C cannot own the house at the same time. Only one of them is supposed to be the new and lawful owner. Hence, the situation is called the double spending problem

The Double Spending Problem

P P SAVANI UNIVERSITY

- Similar to the term blockchain, the term double spending is ambiguous as it is used to refer to the following concepts:
  - A problem caused by copying digital goods
  - A problem that may appear in distributed peer-to-peer systems of ledgers
  - An example of violated integrity in purely distributed peer-to-peer systems

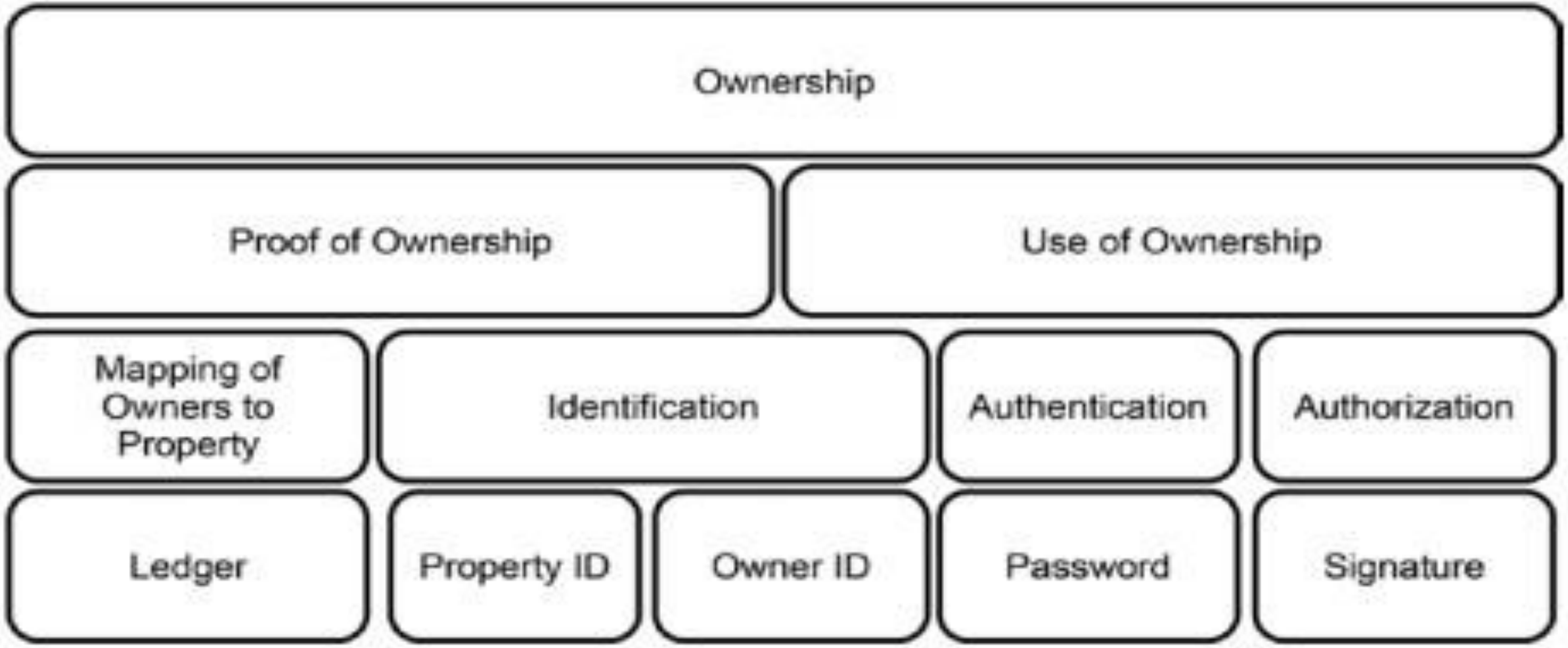
Solving Double Spending as a Problem of Copying Digital Goods

P P SAVANI UNIVERSITY

- The problem of spending digital money or any other digital assets more than once just by copying the data is actually a problem related to the nature of ownership.

- Any accepted means of mapping data that represents digital goods to their owners will solve that problem, regardless of its specific implementation. Even a physical central book or (more realistically) an electronic ledger, regardless of its architecture (centralized or peer-to-peer), can ensure that a digital good will only be spent once, provided the ledger works correctly all the time.

Concepts of Ownership



Concepts of Ownership

P P SAVANI UNIVERSITY

- The concepts on each layer can be seen as realizations of the concepts in layers above them. For example, the proof of ownership requires identification of owners and property alike as well as the mapping between owners and property.
- The use of ownership requires identification as well as authentication and authorization to ensure that only the legitimate person uses the property.
- The boxes in the very bottom row represent the implementation layer.
- They show, for example, that password and signature are concepts used to implement authentication and authorization.
- A ledger can be seen as a concrete implementation of a mapping between owners to their property.

Solving Double Spending as an Example of Violated Integrity in Distributed Peer-to-Peer Systems

- In this context, the architecture of the system is specified but the application domain is left unspecified. Hence, solutions on this level focus on achieving and maintaining integrity in distributed peer-to-peer systems, regardless of their concrete usage.
- However, the concrete usage of a distributed peer-to-peer system determines the meaning of integrity.
- For example, a simple file-sharing application may consider different aspects for defining integrity as compared to a system that manages ownership in a digital currency.
- Hence, the question of whether the blockchain-technology-suite is the right tool for achieving and maintaining system integrity cannot be answered without knowledge of the specific application goals. Hence, it could be possible that in specific application areas of distributed peer-to-peer systems, other technologies, data structures, and algorithms are more suitable for achieving and maintaining integrity.

Hashing Data

P P SAVANI UNIVERSITY

- Hash functions are small computer programs that transform any kind of data into a number of fixed lengths, regardless of the size of the input data.
- Hash functions only accept one piece of data at any given time as input and create a hash value based on the bits and bytes that make up the data. Hash values can have leading zeros in order to provide the required length.
- There are many different hash functions that differ among others with respect to the length of the hash value they produce.
- An important group of hash functions is called cryptographic hash functions, which create digital fingerprints for any kind of data.

Properties of hash values

P P SAVANI UNIVERSITY

- Cryptographic hash functions have the following properties:

1. Providing hash values for any kind of data quickly
2. Being deterministic
3. Being pseudorandom
4. Being one-way functions
5. Being collision resistant

Deterministic

P P SAVANI UNIVERSITY

- Deterministic means that the hash function yields identical hash values for identical input data.

- This means that any observed discrepancies of the hash values of data must be solely caused by the discrepancies of the input data and not by the internals of the hash function.

Pseudorandom

P P SAVANI UNIVERSITY

- Being pseudorandom means that the hash value returned by a hash function changes unpredictably when the input data are changed.
- Even if the input data were changed only a little bit, the resulting hash value will differ unpredictably.
- To put it differently, the hash value of changed data must always be a surprise. It should not be possible to predict the hash value based on the input data.

One-Way Function

P P SAVANI UNIVERSITY

- A one-way function does not provide any way to trace its input values by its outputs.
- Hence, being a one-way function means that it cannot be used the other way around.
- To put it differently, it is impossible to recover the original input data based on the hash value.
- This means that hash values do not tell you anything about the content of the input data in the same way as an isolated fingerprint does not tell you anything about the person whose finger created it.
- One-way functions are also said to be noninvertible.

Collision Resistant

P P SAVANI UNIVERSITY

- A hash function is called collision resistant if it is very hard to find two or more distinct pieces of data for which it yields the identical hash value. Or, to put it differently, if the chance to receive an identical hash value for distinct pieces of data is small, then the hash function is collision resistant.

- In this case, you can consider the hash values created by the hash function as being unique and hence being usable to identify data.

- If you obtained an identical hash value for different pieces of data, you would face a hash collision.

- A hash collision is the digital equivalent to having two people with identical fingerprints. Being collision resistant is mandatory for hash values to be usable as digital fingerprints.

Hash generator

P P SAVANI UNIVERSITY

- http://www.blockchain-basics.com/HashFunctions.html

Tasks to develop blockchain

P P SAVANI UNIVERSITY

- Describing ownership
- Protecting ownership
- Storing transaction data
- Preparing ledgers to be distributed in an untrustworthy environment
- Distributing the ledgers
- Adding new transaction to the ledgers
- Deciding which ledgers represents the truth

Task 1: Describing Ownership

P P SAVANI UNIVERSITY

- Before you can start developing the blockchain, you need to ask yourself what you want to do with it.
- Since you will want to design a software system that manages ownership, you have to decide how to describe ownership first.
- It turns out that transactions are a good way to describe any transfer of ownership, and the complete history of transactions is the key to identifying the current owners.

Task 2: Protecting Ownership

P P SAVANI UNIVERSITY

- Describing ownership by using transactions is just the starting point. Moreover, you need a way to prevent people from accessing the property of others.
- In real life, you can easily prevent people from using your car or from entering your house by using doors with locks.
- It turns out that cryptography provides a way to protect transactions on an individual level, similar to the way doors with locks protect your individual car or house.

Task 3: Storing Transaction Data

P P SAVANI UNIVERSITY

- Describing ownership by means of transactions and having security measures that protect ownership on the level of individual transactions are important steps toward the goal of designing a software system that manages ownership.
- However, you need a way to store the whole history of transactions, as this history is used to clarify ownership.
- Since the transaction history is the core element in clarifying ownership, it must be stored in a secure way.
- It turns out that the blockchain-data-structure is the digital equivalent to a ledger.

Task 4: Preparing Ledgers to Be Distributed in an Untrustworthy Environment

P P SAVANI UNIVERSITY

- The best way to prevent the transaction history from being changed is to make it unchangeable. This means the ledgers and therefore the transaction history cannot be changed once written.
- As a result, you will not have to fear that the ledgers will be tampered with or forged because they cannot be changed in the first place.
- However, having a distributed peer-to-peer system of ledgers that can never be changed sounds like a very secure but pretty useless thing because it will not allow you to add new transactions.
- Hence, the challenge of the blockchain-data- structure is to be unchangeable, on the one hand, while accepting new transactions being added to it, on the other hand.

Task 5: Distributing the Ledgers

P P SAVANI UNIVERSITY

- Once the ledger is append-only, you can create a distributed peer- to-peer system of ledgers by making copies of it available to everyone who asks for it.
-
- However, just providing copies of append-only ledgers does not fulfill your goals.

- A distributed system that manages ownership involves interaction between the peers or nodes, respectively.

Task 6: Adding New Transactions to the Ledgers

P P SAVANI UNIVERSITY

- The distributed peer-to-peer system will consist of members whose computers maintain individual copies of an append-only blockchain-data-structure.
- Since the data structure allows you to add new transaction data, you will have to ensure that only valid and authorized transactions are added.
- It turns out that this is possible by allowing all members of the peer-to-peer system to add new data and additionally turning each member of the peer- to-peer system into supervisors of their peers.

Task 7: Deciding Which Ledgers Represent the Truth

P P SAVANI UNIVERSITY

- Since the transaction history is the basis for identifying lawful owners, having different conflicting transaction histories is a serious threat to the integrity of the system.
- Hence, it is important to find a way either to prevent the emergence of different transaction histories in the first place or to find a way to decide which transaction history represents the truth.
- Due to the nature of a purely distributed peer-to-peer system, the former approach is not possible.
- As a result, you need a criterion for how to find and choose one transaction history that represents the truth. But there is another problem: there is no central authority in a purely distributed peer-to-peer system that can declare which transaction history has to be chosen.

Continue...

P P SAVANI UNIVERSITY

- It turns out that one can solve that problem by making every node in the peer-to-peer system decide on its own which transaction history represents the truth in a way that the majority of the peers independently agree on that decision.

How it works?

P P SAVANI UNIVERSITY

- The procedure to add a new block to the blockchain-data- structure, is not computationally expensive because it only requires adding the hash reference that points to the current head of the chain to the new block header and declaring it as the new head of the chain. The challenge of making the blockchain-data- structure immutable is to make adding a new block a computationally expensive task.

The following aspects need to be considered in the course of achieving this:

- Compulsory data of block headers
- The process of creating a new block header
- Validation rules for block headers

Compulsory Data

P P SAVANI UNIVERSITY

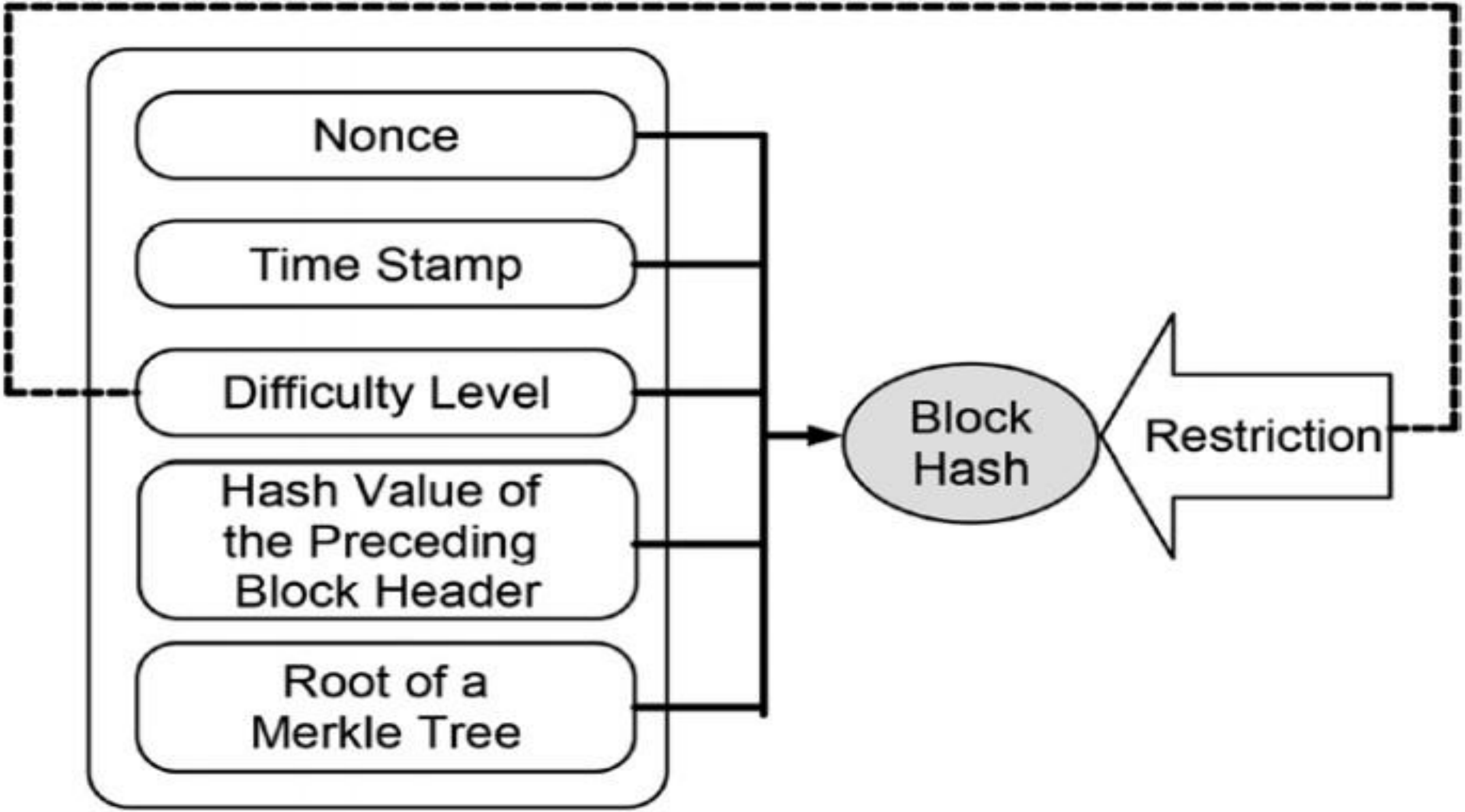
- Every block header of the blockchain-data-structure has to carry at least the following data:
- The root of a Merkle tree containing transaction data
- A hash reference to the header of the preceding block
- The difficulty level of the hash puzzle
- The time when solving the hash puzzle started
- The nonce that solves the hash puzzle

The Process of Creating A New Block

P P SAVANI UNIVERSITY

- Creating a new block involves the following steps:

1. Get the root of the Merkle tree that contains the transaction data to be added.
2. Create a hash reference to the header of that block that will be the predecessor from the new block header's point of view.
3. Obtain the required difficulty level.
4. Get the current time.
5. Create a preliminary block headerthat contains the data mentioned in points 1 to 4.
6. Solve the hash puzzle for the preliminary block header.
7. Finish the new block by addingthe nonce that solves the hash puzzle to the preliminary header.


Nonce
Time Stamp
Difficulty Level
Hash Value of the Preceding Block Header
Root of a Merkle Tree
Block Hash
Restriction

Validation Rules

P P SAVANI UNIVERSITY

- Every block header of has to fulfill the following rules:

1. It must contain a valid hash reference to a previous block.
2. It must contain a valid root of a Merkle tree containing transaction data.
3. It must contain a correct difficulty level.
4. Its time stamp is after the time stamp of its preceding block header.
5. It must contain a nonce.
6. The hash value of all the five pieces of data combined together fulfills
7. the difficulty level.
8. The validation rules ensure that only those blocks are added to the blockchain-data-structure for which the hash puzzle was solved and the computational costs were paid.

Why It Works?

P P SAVANI UNIVERSITY

- The blockchain-data-structure makes any change of its data stand out due to the fragility of the hash references with respect to changes of the data being referred.
- This causes the need to rewrite all blocks that are affected by a manipulation.
- The hash puzzle causes costs for every block that needs to be rewritten in the course of embedding a manipulation.
- The accumulated costs of rewriting the blockchain-data-structure in the course of embedding a manipulation make it unattractive to manipulate the transaction history in the first place. As a result, the blockchain- data-structure becomes an immutable append-only data store.

The Costs of Manipulating the Blockchain Data-Structure

- Let's assume we were going to try to manipulate a particular piece of transaction data that is part of a Merkle tree whose root belongs to a block header located 20 blocks below the current head of the blockchain- data-structure. Embedding the manipulated transaction data requires the following work:
- Rewrite the Merkle tree to which the manipulated transaction belongs.
- Rewrite the block header to which the root of the rewritten Merkle tree belongs.
- Rewrite all succeeding block headers up to the head of the blockchain-data-structure.

The Costs of Manipulating the Blockchain Data-Structure

- Point 2 requires the solution of a hash puzzle because changing the Merkle root changes the hash value of the block header and hence the solution of its hash puzzle. Point 3 requires solving 20 hash puzzles due to successive changes of the hash references to the previous block header.

- Under the assumption that solving a hash puzzle takes on average 10 minutes, we would need in total 210 minutes to embed a manipulation in a transaction that belongs to a block header located 20 blocks below the current head.

- These huge costs deter nodes from changing the blockchain- data-structure.

Is software code mature enough to replace the law?

- In a distributed ledger technology environment, smart contracts are agreed based on a software code and executed on the agreed date (sometimes mercilessly) as the contract itself is the law.
- Although this unalterable nature (or immutability) is the core strength of this technology and enhances trust amongst parties, it also needs to be mature enough to replace the law.
- There have been instances in the past when some of the well-known DLTs had to be "hard forked" – a phenomenon whereby the governing code has to be replaced with a new one.
- In 2016, for example, Ethereum had to be hard forked after long debate amongst the community as an unexpected code path allowed users to withdraw funds and an unknown user managed to withdraw USD 50 million. Not all in the community agreed with the decision, which led to different versions of Ethereum, viz. Ethereum and Ethereum Classic.

Standards are underdeveloped and not mature yet

P P SAVANI UNIVERSITY

- Being at a stage of rapid technological development, there are no mature standards addressing distributed ledger technology yet.
- At this point, there are various competing proprietary and community- managed platforms and frameworks.
- The absence of international standards carries risks related to customer lock-in, lack of interoperability, privacy and security.
- There are international efforts ongoing in these areas, including ISO Technical Committee 307 on Blockchain and Distributed Ledger Technologies and work in ITU's standardization sector ITU-T.

Energy requirement can be high

P P SAVANI UNIVERSITY

- A methodology to build consensus for entering a new data block amongst participating nodes is a core feature of blockchain. There exist several possible ways of reaching consensus, each with its own advantages and disadvantages.
- The one that is employed by Bitcoin and Ethereum, the most famous of blockchain implementations, is proof-of-work (PoW).
- It works on the principle of “hard to create, easy to verify”, which means lot of energy needs to be spent by the node to earn incentive tokens. For a large chain like Bitcoin, estimates suggest data size exceeding 100 gigabytes and electricity requirements more than the entire country of Ireland.
- Although this is true for the PoW methodology, other alternatives such as proof-of-stake (PoS), Byzantine fault tolerance algorithm, and delegated proof-of-stake model require less energy. However, they come with their own disadvantages, for example in the case of PoS, users with more stakes will have greater control on decision-making.

Trusting the blockchain developers and managers

- A very high level of trust is placed on the developers and managers of the blockchain.
- It is a new technology where a large number of entities are innovating to create solutions.
- The focuses, owners and software implementations vary.
- Implementations of these technologies are largely dependent on the community of developers backing the project or the owner.
- A decision to soft fork or hard fork a project, or to change the cryptography algorithm, will be driven by the nodes and participants in the blockchain.
- These decisions are driven by codes that govern the consensus and the
- community developing it.
- At the same time, it is important to build resilience into the networks so that they can be entrusted with critical data, information and services.
- Carrying out a risk assessment of the project is important before making a choice.

Increased responsibility on the user

P P SAVANI UNIVERSITY

- By its very design, blockchain implementation does not have a central authority – at least in the case of public blockchains such as Bitcoin – which puts additional responsibility on the user.
- There is no entity to go to in the event of individuals losing private keys (or incurring losses as a result of revealing a private key).
- Also, there is no feature to restore forgotten passwords and usernames that individuals are used to.
- Individuals need to exercise great caution, just as on the Internet, before publishing anything. The importance of entering the correct data is very important too as it is very difficult to make corrections later.

Implementing data privacy legislation

P P SAVANI UNIVERSITY

- Data protection and privacy is a major concern and initiatives to prevent their abuse are being taken by countries and regions (e.g. Association of Southeast Asian Nations (ASEAN), European Union General Data Protection Regulation (EU GDPR).

- For example, the EU GDPR has instituted the “right to forget” whereas the design of DLTs is oriented towards “never to forget”. Although there is a possibility of keeping identification unknown in the system, it raises security concerns largely in relation to anti- money laundering (AML) activities and know your customer (KYC) requirements.

Policy and regulatory risks

P P SAVANI UNIVERSITY

- The policy and regulatory framework around blockchain is in its infancy and therefore entails high risks.
- The fluctuations in the price of Bitcoin andthe reports of hacking of cryptocurrency have resulted in increased regulation by a number of countries and has attracted regulatory interest.
- These regulations vary from a complete ban on holding cryptocurrency (e.g. Bangladesh), a ban or regulation on cryptocurrency trading (China,Saudi Arabia) to a ban on holding initial coin offerings (ICOs).
- A number of blockchain projects, especially those dealing with currency or cross-border transactions, requires KYC compliance and it is important to understand the national framework before delving into these projects.
- At the same time, governments see DLTs as a high potential technology and are investing in the use of its application. A project without the use of cryptocurrency in general will have less regulatory challenges than those with it.

Speed of transactions

P P SAVANI UNIVERSITY

- The speed of transaction is an important element as some of the public blockchains do not have high transaction speeds. On Bitcoin blockchain, a new block emerges on average every ten minutes but is not guaranteed; and this block time is different for every blockchain.

- For scalability, it is important to understand the requirement of applications in terms of speed (transactions per second (tps)) before choosing a solution. Theoretically, Visa network can handle about 50,000 tps, which is a lot more than is offered by most mature blockchains today.

Malicious users

P P SAVANI UNIVERSITY

- In the absence of identification of a third party, the system is prone to risks from malicious users in systems that are pseudonymous, that is with no requirement to disclose identity.

- Although DLTs are designed to disincentivize malicious intent, there can be situations where malicious users have greater incentives to game the system and at least cause harm in the short term and may call for a hard fork. These situations are more likely where they gain greater control of the system.

Identity and security

P P SAVANI UNIVERSITY

- Public blockchains carry out transactions based on the public and private key of the individual and do not keep the mapping of the identity with the key.
- This raises security constraints for the law enforcers and applications where identity is important.
- In contrast, there are privacy concerns in disclosing identity on permission less blockchains that require data to be public facing and transaction histories to be disclosed.
- Most DLTs use encryption algorithms that are hard to break by normal non-quantum computers.
- Going forward, where quantum computing (relying on cubits rather than bits) gains momentum and enhances computing powers, these encryptions are not secure enough.

Continue…

P P SAVANI UNIVERSITY

- There have been a large number of successful attacks on DLTs and there are security risks associated with DLTs30 (e.g. blockchain attacks, phishing, malware, cryptojacking, endpoint miners, implementation vulnerabilities, wallet theft, technology attacks, legacy attacks which have been modernized, dictionary attacks, quantum computing-based attacks).