SECTION 2

DISTRIBUTED LEDGER AND BLOCKCHAIN FUNDAMENTALS

This section will discuss the technical ideas underpinning distributed ledger and blockchain.

We will introduce distributed ledgers and what the crypto, or cryptography, in cryptocurrency means, including a few key cryptographic concepts. We will also cover how a blockchain is structured and why it is useful.
In order to understand a distributed ledger, we first need to understand what is meant by a centralised ledger. Let us remind ourselves of the purpose of a ledger.

Essentially, a ledger is a book or file for recording monetary transactions, with debits and credits recording movements in separate columns. A ledger begins and ends with a monetary value.

2.1 DISTRIBUTED LEDGER

What is a distributed ledger, and why is it important for more than just Bitcoin?

In this context we'll be talking about a ledger as a store of data, or a **database**. In Bitcoin, it is a store of transactions, but it could also be other types of records.

Today, most data stored in an organisation will be contained within relational databases, big data solutions, data lakes, or equivalents. It is all managed within the boundaries of an organisation, and not designed to be shared. If data needs to be shared then it inevitably becomes difficult to do so. You need to solve problems, such as who owns the data, and who can access and update it. Even if you are able to share, sharing data between two parties is very different to sharing data between thousands!

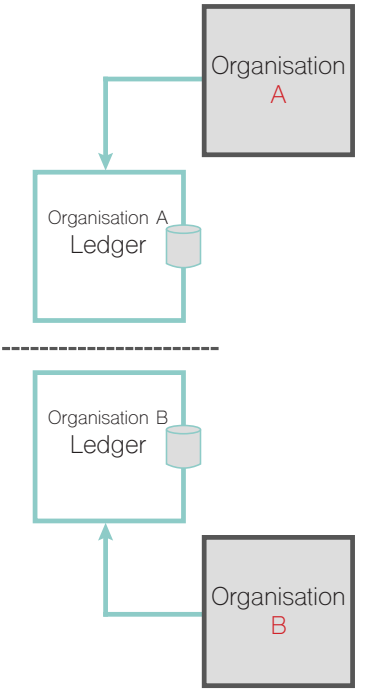
Data held within most organisations is fragmented because there isn't necessarily a single source of the truth. The data could be split between several databases, or replicated in a way that makes consistency difficult and multiple reconciliations necessary.

So next we will look at different ways of sharing this data:

- Separate ledgers with reconciliation
- Centralised ledgers
- Distributed ledgers

SEPARATE LEDGERS

Separate Ledgers are where two parties maintain their own ledgers, each responsible for their own information.



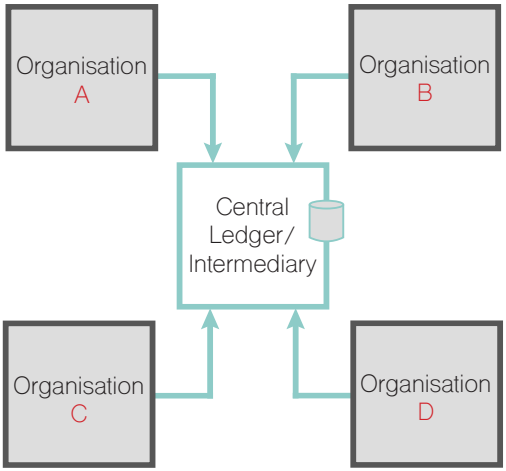
It is the responsibility of the organisations to maintain their own data, so there need to be processes and reconciliations put in place in order to ensure that the two ledgers are synchronised. This must be managed very carefully, or else the ledgers could drift apart and become inaccurate. If such a problem is found then it may be unclear which party has the accurate record, who is responsible for correcting the issue, and in which ledger this should be recorded.

And it is not just data that needs to be synchronised. Permissions and validation could be different in different ledgers, meaning one ledger could deem entries to be valid that are not valid elsewhere.

And if more parties become involved then the amount of communication channels increases, meaning that reconciliations become a lot more complicated.

CENTRALISED LEDGER

A Centralised Ledger involves the main store of data being held in a single place by a central authority or intermediary who is responsible for managing this data.

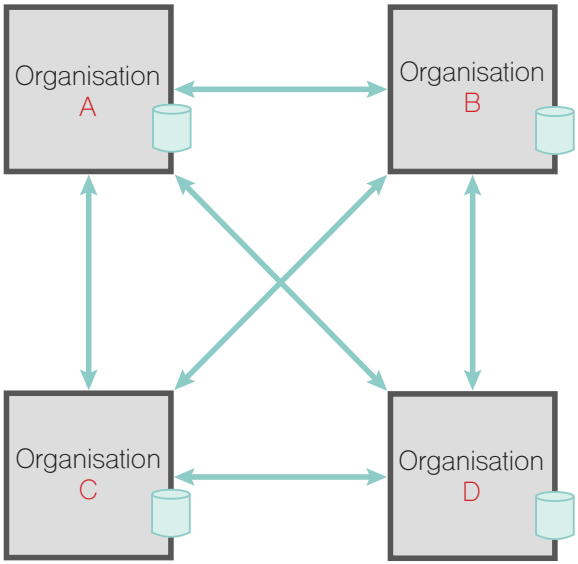


A centralised ledger solves the problem of reconciliation because everyone is using the same ledger.

The downside of this approach is that the central authority must manage the interactions, which means that the organisations are placing trust in a single external third party. The rules of the interactions must also be clearly defined; for example, permissioning and data validation.

DISTRIBUTED LEDGER

A Distributed Ledger allows a ledger to be shared without needing a central authority. This solves the problems presented by separate ledgers and centralised ledgers.



It is a database that is **decentralised** – i.e. distributed among multiple participants and/or locations. It is a solution to the ownership, permissions, sharing, and reconciliation problems that are inherent in a centralised ledger. All participants work together to store, distribute, and validate data. In the world of blockchain, we refer to these individual databases as **NODES**.

Bitcoin operates on the bitcoin ledger which is decentralised, meaning that there is no central authority sitting in the middle managing transactions. It was indeed designed to avoid having a centralised ledger, and a centralised authority.

2.2 CRYPTOGRAPHY

Two important concepts used by Distributed Ledgers are **cryptography** and **consensus mechanisms**. Cryptography is used to ensure the integrity of all data. Consensus mechanisms are used to ensure all participants in the distributed ledger agree on a single version of the truth.

Cryptography is the practice of securing data.

Cryptocurrencies, blockchains, and DLT use three important cryptographic concepts:

- Cryptographic hashing.
- Public key encryption.
- Cryptographic authentication, or digital signatures.

Modern cryptography is a weighty subject, based in mathematical theory. You do not need to understand the technical details of how it works, but rather what it does and why it is used.

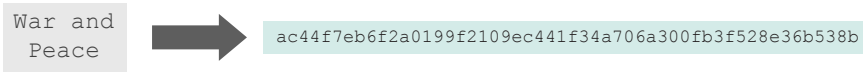
The cryptography used in blockchain is thoroughly proven. The same concepts underpin e-commerce and almost all kinds of digital 'secret' keeping.

Modern cryptography is based on improvements in computing power, and breakthroughs in number theory in the mid-1970s.

2.3 CRYPTOGRAPHIC HASHING

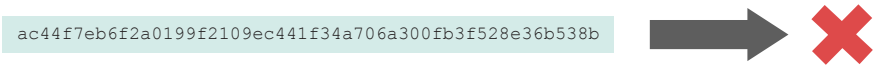
Cryptographic hashing is a function, meaning a formula that, when given an input, gives an output called a 'hash value'. It takes any input data of arbitrary length, and returns a short, fixed length value that uniquely represents the input data.

Hashing is used to verify data. Changing the input data even slightly will cause the hash value to change. A hash value will always be short, so it is an efficient way of verifying data.



It stores enough information to uniquely identify the data, but not the content itself. It is akin to a digital fingerprint.

Hash functions are one way, or **asymmetric**. This means that the same hash value can always be encoded from the same input data, but you can never take a hash value and derive what the input data was from this.



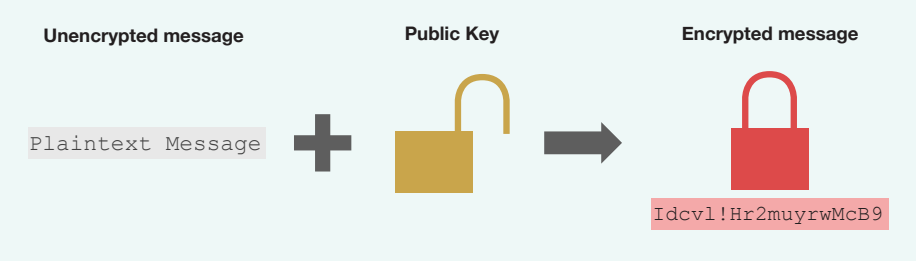
A lovely way of explaining this is used by Adrian Patten, the founder of Cobalt, a DLT-inspired financial services business. He states that it's like putting a piece of steak into a mincer – whatever you do, you can't reverse the process and recreate the steak! In the same way, you can't recreate the input data from the cryptographic hash.
Hashing is performed with hash algorithms. If you use the same hash algorithm, whether implemented on a PC, in a browser, smartphone, or tablet...with the same input data, you will always get the same hash value. Examples of hash algorithms are SHA-256 used by Bitcoin, or KECCAK-256 used by Ethereum. Generally, they are actually represented as alpha-numeric characters of 27 to 34 digits in length.

2.4 PUBLIC KEY CRYPTOGRAPHY

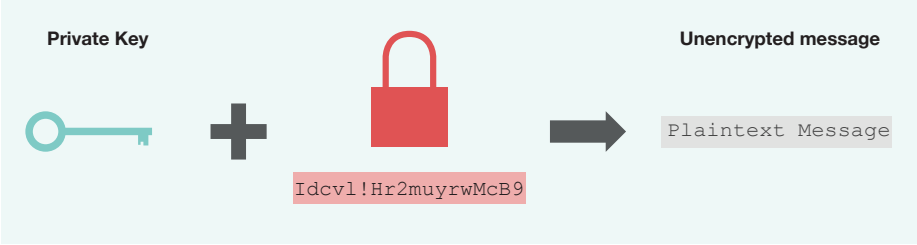
Public key cryptography is used for encryption and authentication. It is a way of keeping data secret. A **key** in cryptographic terms is the information you need to translate encoded cyphertext to readable plaintext.

Public Key Cryptography was invented to solve the problem of key distribution. Symmetric encryption means the same key is used to encrypt and decrypt. If you are going to pass a secret to someone, you must both have the key, and both keep it secret. This is a logistical problem because you must manage separate keys with everyone that you communicate with.

Public key cryptography is asymmetric. It uses two different (but linked) keys. There is a public key for encryption, so anyone can encrypt a message, and a private key for decryption, so only the private key holder can decrypt the message. This solves the key distribution problem. The private key holder can distribute the same public key to everyone, safe in the knowledge that they can send encrypted messages, but they are never able to read other messages to the private key holder.



A message is encrypted with the public key, resulting in an encrypted message



The message can be decrypted using the private key to get back to the unencrypted message

With Bitcoin, your private keys live in your wallet (see chapter 7). This is the information only you hold, which ensures that your bitcoins are safe.

The public key is akin to an open padlock. After encryption, the padlock is closed, and a message locked inside. Only the private key can unlock it.

Public key cryptography is based on mathematical problems that are easy to solve in one way, but incredibly difficult to the point of being virtually impossible to solve the other way around. They are described as trapdoor functions. You can go one way through the trapdoor, but it springs shut, and without a key to unlock it you can't get back through.

There are several mathematical 'problems' that are used. One is based on prime factorisation, where two large prime numbers are multiplied together to get an even bigger semi-prime number. The two individual prime numbers (prime factors) are the private key, and the product of them is the public key. It is virtually impossible to know how the product was made without knowing one of the prime factors.

Bitcoin uses a different technique called Elliptical Curve Cryptography, based on something called the elliptic curve discrete logarithm problem. That's all you need to know for now. Too much more explanation would be too much for this book!

2.5 AUTHENTICATION AND DIGITAL SIGNATURES

Digital signatures are used to validate data to prove its authenticity and integrity. They definitively tell you who authored the data in question, and confirm that it has not been changed since the signature was created. They are effectively a stamp of authenticity, a wax seal proving the provenance of the information. You cannot change the content of the message without damaging the seal.

A digital signature can be used to authenticate an email, so you can be certain that the email from your bank really did originate from your bank, or to authenticate an invoice so you know that the account number is correct, and that a fraudster hasn't sent you an amended version with their own account details.

In Bitcoin, digital signatures are used to prove that you are the instigator of a transaction, and also to authenticate the details, the amount, and the recipient of that transaction.

Digital signatures are based on hashing and public key cryptography. Public keys can work in two ways. They can be used for one-way encryption, as we have seen, but they can also be used the other way around for one-way decryption. This effectively means that anyone can decrypt, but only the private key holder can encrypt.

Digital signatures use private keys for encryption and public keys for decryption. The private key holder is the only person who can create a suitable digital signature. They prove the authenticity and integrity of the data. Public key holders can confirm and validate by using the public key, and decrypt the signature safe in the knowledge that the only person who could produce a correct digital signature is the private key holder.

The process is as follows:

SIGNING

Step 1 - Data that is going to be signed is hashed

Step 2 – The hash is encrypted with the private key

The message is now signed with a digital signature

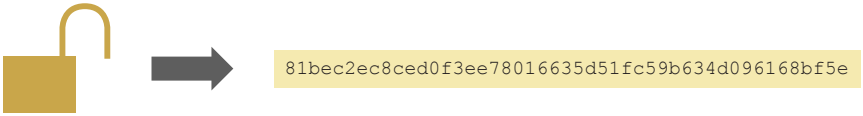
Message to be signed

de2959c902d7375a489d976f4a373e19204a065

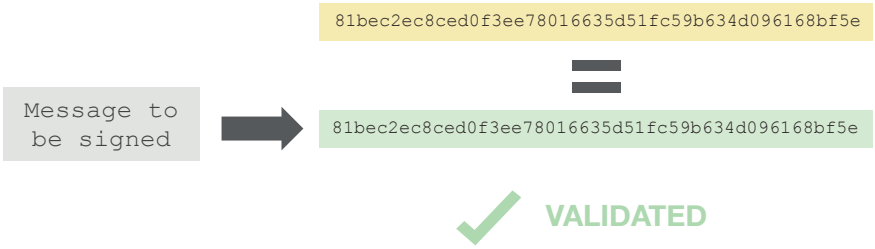
VALIDATING

Step 1 – The signature is decrypted with the sender's public key giving the hash

Step 2- The decrypted hash is compared to our own hash of the data



Step 3 – If the hashes match we know the message is valid and the sender is authentic as they are the only key holder able to produce a correct hash. If the message were changed the hash comparison would flag this up.



2.6 BLOCKCHAIN

A blockchain is a type of database, which is simply a collection of organised data. As the name suggests, a blockchain is a chain of blocks, with each of the blocks containing data. In Bitcoin, these are batches of transactions, but the data could be literally anything. In order to achieve scale and speed, many solutions just distribute the cryptographic hash. The data then never leaves the home environment, which is important for security.

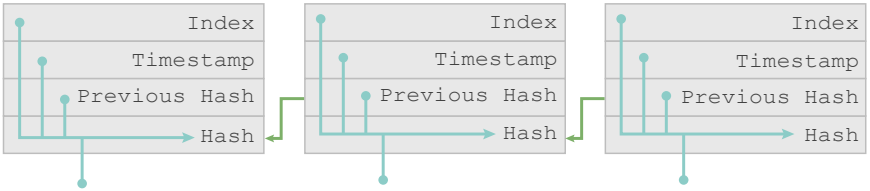
Blockchains are often described as **immutable**. This means that once data is written to the blockchain it cannot be erased or modified. This is because blockchains add data sequentially, whereby new blocks are added onto the end of the chain. As we will see, the cryptography used ensures this immutability. Because blockchains are distributed, and we do not necessarily know or trust everyone else in the blockchain network, we need these immutable properties as part of the architecture in order to ensure confidence in the data.

Immutability doesn't mean that data is impossible to change, though. As with a ledger you can make amendments through journals, by adding new records to reflect how the data has changed. Hashing is used to ensure the immutability of the ledger, and digital signatures are used to ensure the provenance and identity of the data. If data is sensitive, public key cryptography can be used in order to encrypt the data.

2.7 BLOCKCHAIN STRUCTURE

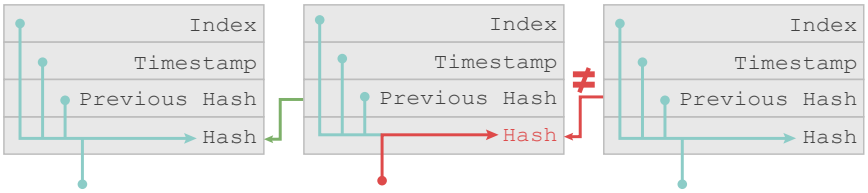
Blocks are chained together in a blockchain. A block will have a header comprising information about the block and data, and a list of records and transactions.

The header will contain the index of the block, a timestamp for when the block was created, the hash of the previous block in the chain, and its own hash, including the header and data.



If any part of the header or the data changes then the block hash will also change. A block's hash will include the previous block's hash, so if this changes, or any previous block changes, then there will be an avalanche effect on subsequent blocks' hashes.

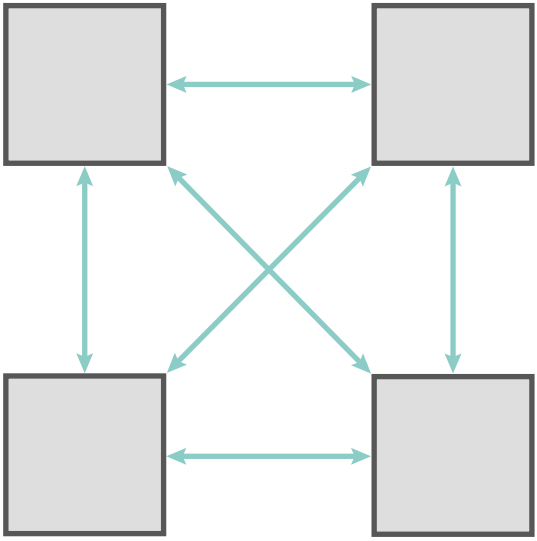
This is how the blockchain is secured. The cryptographic hashing ensures that you can't change blocks in the chain. If you do then the change will be invalid.



In this example, the data in Block 1 is amended. This changes the hash of Block 1. The previous hash from the next Block 2 no longer matches the hash from Block 1.
The amendment is rejected because it is invalid.

2.8 CONSENSUS MECHANISM

There is no central authority in a distributed ledger, so there is no central entity that makes decisions on whether data is valid. Instead, decisions to add transactions to the distributed ledger are made on a **consensus** basis.



Using a **consensus mechanism**, the blockchain network self-governs and prevents malicious or erroneous transactions.

Different blockchain implementations use different consensus mechanisms. The consensus mechanism also depends on the nature of the network, and the level of trust involved. Whether a network is completely open to the public, or in a private network – for example, in an enterprise environment – will then influence the mechanism used.

The consensus mechanism chosen also affects the speed that blocks are added to the chain. Here again, the usage will dictate the type of mechanism used. In an enterprise environment,

speed could be of primary importance, but in a public blockchain security and integrity are foremost.

Dan Klein, Chief Data Officer at Voltech, an amazing data scientist and engineer, describes this consensus mechanism as akin to out-sourcing the audit function.

Bitcoin also uses a Proof of Work (PoW) mechanism. This is relatively slow, and purposely so. Blocks can only be added to a chain, or mined, after a computationally expensive puzzle has been solved that is linked to the hash of the block. This stops anyone maliciously amending the blockchain. Rewinding, replaying the transactions, and correcting the hash chain would represent a huge amount of work. Furthermore, the amended blockchain would never catch up with the legitimate blockchain, due to the time and computational expense involved in the re-calculation.

In Bitcoin, whomever solves the puzzle first gets to add the block to the blockchain, and is rewarded with 1 bitcoin.

2.9 TYPES OF BLOCKCHAIN

There are different types of blockchains for different use-cases**:**

PUBLIC

Public blockchains, such as Ethereum or Bitcoin, live on the internet and are completely decentralised. Anyone can join the network, and they are what is known as 'low trust', meaning that other nodes aren't trusted. This means that a lot of emphasis is therefore put in place in order to ensure the integrity of the blockchain. The low trust in the individual nodes is compensated

for by the high trust in the process.
Any user can read or write to the blockchain, assuming that the transactions are valid.

PRIVATE

A private blockchain is controlled by a single organisation, and permissions to read or write are tightly controlled. It could be on the internet with restricted access, or it could be inside a corporate network with no access to the outside world.

This is in opposition to a public blockchain. In a private blockchain individual nodes would be highly trusted, meaning there can be less emphasis on the security mechanisms, and focus could instead be placed on performance.

CONSORTIUM

A consortium blockchain is a partially private blockchain that is open to select members instead of being open to the public. This enables organisations to collaborate with their data in specific ways. The consortium would be organisations in similar industries, e.g. banks, trying to solve a common problem.

ENTERPRISE BLOCKCHAIN

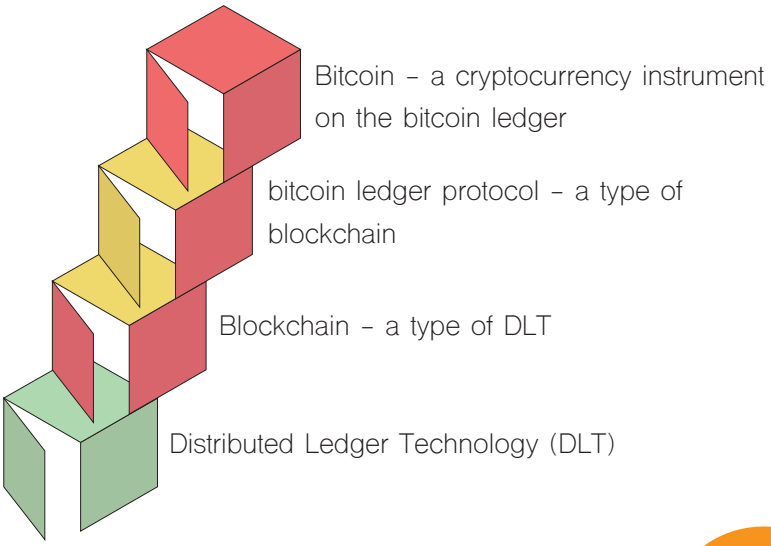
Early on, companies tried to co-opt the existing public blockchain technologies and use them for business purposes. But they found them lacking in certain areas, including data privacy and performance. Since then, 'enterprise' blockchains have begun to appear that are specifically built for businesses. These include IBM Hyperledger, Corda R3 and JP Morgan's Quorum.

In particular, R3's Corda ledger also solves for a regulatory sensitivity issue: to achieve legal finality, consensus just isn't good enough. The protocol for adding to the ledger has to be legally binding, final and absolute.

DATA SECURITY AND BLOCKCHAINS

Data can be secured on a blockchain using public key cryptography in order to keep it secret. If data is especially sensitive, or the size of it is particularly large, then it can be stored separately to the blockchain, and referenced via a cryptographic hash or digital signature.

This is known as keeping the data "**off chain**". The signature or hash uniquely identifies the data, so there is still immutability, and it is also guaranteed that the data has not changed. Furthermore, many scale challenges are addressed by reducing the actual data that is distributed – to just the hash!



Bitcoin Use-case

Simple visual above – the point being that the Bitcoin is a cryptocurrency on the bitcoin protocol blockchain. See later chapters. It is just ONE use-case.

SUMMARY

- ✓ **Blockchain** is where data is organised into encrypted blocks and chained together.
- ✓ **Distributed Ledger** is a decentralised database where multiple participants store, distribute and validate data.
- ✓ **Cryptography** is the practice of securing data with very complicated mathematical equations.
- ✓ **Cryptographic Hashing** is a formula that converts data into a unique short hash value. If that data is changed in any way the output hash will change.
- ✓ **Public and Private Key Cryptography** is used for authentication – the public key encrypts the message but only the private key holder can decrypt.
- ✓ **Digital Signatures** authenticate the provenance of data and definitively tell you who authored.
- ✓ **Immutable** means that once written to the blockchain the original data cannot be erased or modified. Amendments are only possible through new records.
- ✓ **Consensus Mechanism** is a self-governing process by which decisions are made to add data to the distributed ledger.