United Nations ECE/TRADE/C/CEFACT/2018/9

Economic and Social Council

Distr.: General
20 April 2018

English only

Economic Commission for Europe

Executive Committee

Centre for Trade Facilitation and Electronic Business
Twenty-fourth session
Geneva, 30 April and 1 May 2018
Item 7b of the provisional agenda
Recommendations and standards: Other deliverables for noting

White Paper on technical application of Blockchain to United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT) deliverables – Annex

An Introduction to Blockchain

Summary

Blockchain applications are providing new ways of exchanging data in a secure manner. Many UN/CEFACT domains have expressed an interest in applying blockchain technology to the standards they develop. In 2017, the Bureau decided to launch a single blockchain project in order to provide a harmonized framework on how each project team should approach Blockchain. The result was a project that aims to develop two White Papers. The first on the possible impact of blockchain technologies on UN/CEFACT deliverables, the second on the potential use cases of blockchain in the supply chain and beyond.

Both white papers are currently under development. This current document is Annex of the first white paper and is an explanation of blockchain technology for lay persons. The remainder of the draft of the first white paper will be presented as an informal document (ECE/TRADE/C/CEFACT/2018/INF.1).

This document is presented to the 24th session of the Plenary for information.

Annex

I. Blockchain - How it works

1. At its heart, blockchain is a cryptographic protocol that allows separate parties to have shared trust in a transaction because the ledger cannot be easily falsified (i.e. once data is written it cannot be changed). This security is due to a combination of factors including the cryptography used in a blockchain, its consensus/validation mechanism and its distributed nature.

2. This annex does not aim to provide an in-depth review of blockchain technology - there are plenty of web resources to help readers achieve that goal. Rather, it will cover the core concepts which are needed to understand the potential application of blockchain in international supply chains.

3. First, some nomenclature:

- **Node**: System that hosts a full copy of the blockchain ledger.
- **On-chain transaction**: Automated procedure that creates or updates the state of an address in the blockchain database by appending new data to the ledger. Examples include digital asset exchange, or execution of an automated business process.
- **Validation**: Work performed by all nodes in parallel, that verifies transactions using a consensus algorithm. Different networks may use different consensus algorithms. When mutual validation results in a consensus, then the nodes all commit (record) the transaction onto their blockchain.
- **Block**: Data that is appended to the ledger by consensus. Once a block is written to the chain, it cannot be changed or deleted (without replacing all subsequent blocks).
- **Hash**: Fixed size, unique cryptographic fingerprint of data. A hash is a one-way function; this means that given the data, one can easily verify that the hash is the correct one for that data. However, it is not possible to reverse-engineer the hash, so you cannot use it to re-create the data. This is a key feature because it allows users to confirm that no changes have been made. For example, even an additional space or empty line in a text would change its hash.

4. An important characteristic of blockchain systems is the way consensus allows users to trust that transactions have been executed and trust information about those transactions (for example, their date and content). As a result, blockchain systems can be used as an independent umpire in processes that might otherwise expose participants to the risk of one party not living up to its contractual obligations (counterparty risk) and third party guarantors are reluctant to intervene and assume part of that risk. In the case of public blockchains, the umpire is the society of all nodes that choose to participate in the consensus. In the case of private blockchains, the umpire is the consortium of nodes trusted to (given permission to) create consensus on the network.

A. It’s a distributed ledger

5. Ledgers are a kind of database, kept digitally or with paper records, where transactions are recorded once and not subsequently updated (also known as a journal database). Each record can be read many times but written only once. The term ledger comes from accounting where entries, once written into a ledger (accounting journal), cannot be changed.

6. A blockchain is described as distributed because there are multiple copies which are kept on different nodes. The multiple copies are updated in a coordinated way that ensures they remain consistent, using a consensus algorithm (of which there are many). Specifically, the consensus algorithm decides (by mutual agreement between the nodes) which block is added to the chain next. In essence, a blockchain database is a sequence of data blocks that have been added in a specific order, by consensus of the network operators, to each of multiple copies of the ledger and where each block contains a fingerprint (hash) that can be used to verify the content of all the previous blocks.

B. It writes transactions

7. Each block of data written to the ledger contains at least one or many records of transactions. A familiar example of a transaction would be "debit one coin from account A, and credit one coin to account B", although many other kinds of transactions are possible. Some blockchains support a limited sub-set of transactions (operations or algorithms), such as this simple double-entry bookkeeping operation. Some blockchains support a much wider set of transactions covering any solvable algorithm (i.e. a Turing-complete computer programming language[1]). These types of transactions are variously called smart-contracts, chain-code, transaction families, or other, equivalent terms. In summary, all blockchains support a variety of data operations on their chains, but not all blockchains support Turing-complete transaction languages.

C. To a cryptographically signed block

8. Blockchains implement two kinds of cryptographic technology: hash functions and public/private key cryptography. Hash functions are used to construct the fundamental proof that links each block to the rest of the chain before it. Hashes, in a different context, can also be used to provide proof of validity for data that is referenced by blocks. Public/private key cryptography is used for identifying transactors and controlling access to data. An analogy is e-mail where the public key is your email address (which others can use for sending messages to you) and the private key is your password which gives access to the private material which is your messages. So, on a blockchain, a public key can be used, for example, to implement a transaction that sends a document or a payment to a party, but only the party with the private key can access those documents or payments after they are sent.

D. That independent nodes must verify

9. There are various consensus algorithms used by different blockchain systems. For example, Bitcoin uses proof of work algorithms which allow miners to recover the cost of computationally expensive work in exchange for transaction fees. Permissioned ledgers use a consortium of collectively trusted (but not necessarily individually trusted) nodes to agree on the output of a consensus process, which are generally cheaper and faster than Bitcoin's proof of work. All consensus processes require a mechanism to settle disputes, or

[1] Turing complete programming language is capable of solving any mathematical problem computationally (if you know how to program it). In general, this means it must be able to implement a conditional repetition or conditional jump (while, for, if and goto) and include a way to read and write to some storage mechanism (variables).

uncertainty, about which block should be written next. Most of these mechanisms are based upon using the block which is agreed upon by more than 50% of the nodes.

10. The nature of the consensus mechanism determines some key characteristics of a blockchain system. For example, Bitcoin has deliberately made mining (the creation of blocks) expensive. This protects the blockchain by making the cost of capturing more than 50% of the nodes (the number needed to approve a block, and thus to manipulate the blockchain) prohibitively expensive. To compensate for this cost, miners are rewarded both an amount of Bitcoin for each block they create and fees for each transaction written to the blockchain[2]. Each block has a size limit and transaction costs are determined on a free market basis, so the more transactions are requested, the more the price increases for each transaction. This is necessary for the Bitcoin economic operating model, which seeks to obtain an honest consensus in an unregulated market of potentially anonymous and economically rational operators (i.e. operators who might, being anonymous, and having no costs for doing so, steal assets). As an additional incentive, if a node/miner does not accept the block voted on by over 50% of the other nodes, it is, effectively, kicked off the blockchain, thus losing the possibility of earning future Bitcoins and transaction fees. As a consequence, Bitcoin has extremely low bandwidth (due to the cost of generating blocks) with transactions taking more than 10 minutes to be confirmed. In addition, its very large number of nodes and users (generating large amounts of data), together with its block size limits, makes storing data on the Bitcoin blockchain expensive as well as being inefficient (given the duplication of information across all nodes, it is generally inefficient to store significant amounts of data on any public blockchain). Bitcoin still supports many billions of US dollars worth of Bitcoin and other high-value transactions, but its speed and volume limitations make this blockchain unsuitable for many enterprise applications.

11. Permissioned ledgers strike a different balance between bandwidth, capacity and trust. For example, because they have more control over who participates, permissioned ledgers can use other consensus mechanisms, even if some of them are somewhat less robust than the proof of work used by Bitcoin. For examples, there are consensus mechanisms based on the amount a node has invested in a network (called proof of stake), or where a consensus by a subset of nodes is verified by a larger group. In addition, there is a great deal of research going on to identify and test a range of other consensus mechanisms. Using these alternative consensus mechanisms, some permissioned ledgers can support hundreds or even thousands of transactions per second (rather than an average of one new block per 10 minutes, as with Bitcoin) and petabyte scale databases.

E. The block is written to the ledger after it is verified

12. When consensus is reached (which includes agreeing that a block contains legitimate data, and that it is the block that should be written next), each node adds the agreed block to their local copy of the ledger. In this way, all nodes maintain an identical copy of the ledger each time a block is written. This is guaranteed (proven) by the next block to be written, because it will contain a hash of the block before it.

F. The new block is linked to previous blocks - creating immutability

13. Recall that a hash is a one-way function that produces a unique fingerprint of some data. Also note that a hash function produces a fixed-size fingerprint regardless of the

[2] Bitcoin is designed so that, over time, mining rewards are reduced with the objective of eventually having all mining rewards come from transaction fees.

amount of data being hashed. For example, there is no way to know from looking at the hash if the data was a single small document or a database holding many billions of records.

14. Each block in a blockchain contains some transaction data, plus the hash of the previous block (which is always the same size, no matter how much data it represents). Given a consensus that this new block forms part of the chain, it is possible to verify the previous block from its hash. And from the previous block, the block before it, and so on all the way to the first (or genesis) block in the chain. The hash of the previous block is said to be anchored in the subsequent block.

15. Tampering with the contents of any block in the chain will change the hash of that block, which will change the hash of the block after it, and so on for every subsequent block in the chain. If this occurs then the tampering is easily detectable by any node, and the consensus algorithms will prevent new blocks from being written to a chain because the hashes don’t match.

16. This characteristic is the origin of the word “chain” in “blockchain” because each block is anchored to the previous block and proves the existence of all the data it references going back to the first “block” of data in the “chain”.

II. Blockchain - Types

A. Public Ledgers

17. Public ledgers can be read by anyone. They are also permissionless in the sense that anyone can participate and utilise consensus mechanisms without depending on a regulator to enforce acceptable behavior. Bitcoin, Ethereum and more than 10 other cryptocurrencies with market capitalization over USD 1B operate this way, allowing any transaction that is logically valid even between anonymous parties.

18. One of the fears about blockchain technology is that, if a malevolent actor were to control a majority of the nodes, then they could decide to reach a consensus in contradiction of the interests of other stakeholders. This threat is described as a Sybil attack in the cryptographic literature. A successful Sybil attack on a public blockchain cryptocurrency could result in a catastrophic redistribution of assets or double spending. Public ledgers are designed to operate according to rules that do not require governance or regulatory mechanisms to intervene in order to prevent antisocial transactions, because those mechanisms might themselves be exploited for antisocial outcomes, for example, if they were to be hacked by a third party or abused by the trusted regulators. These systems operate with absolute trust in their algorithms and are designed to avoid any need to trust any counterparties. This is why (public) blockchains are sometimes referred to as being trust-less.

19. Public ledgers typically compromise other aspects of performance in order to achieve strong resistance to Sybil attacks. They also rely on the transparency of the public ledger, and also on the transparency of the open source software involved.

B. Permissioned/Private ledgers

20. Like conventional (operational/analytic) databases, the contents of a private blockchain ledger may be a guarded secret that is only available to selected users (and node operators) through a role-based access control mechanism. Unlike a traditional database, a

private blockchain ledger is immutable (cannot be updated) and transactions are verified by a consensus mechanism that is established by the network operators.

21. Private ledger technology is typically applied in enterprise use-cases where immutable transactions are required, that can be verified by a closed community of nodes. These nodes may be independent of parties to the transactions on the blockchain and may be subject to oversight and governance that is not possible (or considered desirable) in a permission-less blockchain system.

22. Permissioned ledgers operate with a different threat model to the public ledgers. The operators of permissioned ledgers are not anonymous, they are subject to some kind of governance controls and are collectively trusted by the users. Antisocial behaviour of a node or participant could result in that party being evicted from the network, and their transactions blocked or even rolled-back from the blockchain by consensus of the remaining operators. The expectation of users of a permissioned ledger is that the operators will intervene in antisocial behaviour but not commit antisocial behaviour themselves.

23. On permissioned ledgers, the level of security, and so the confidence users can have in the immutability of the data, varies depending upon the rules established for that permissioned ledger (including its consensus mechanism). Permissioned ledgers can also create a false sense of security because only trusted participants are allowed to maintain nodes and participate in verification. At the same time, even trusted participants can become untrustworthy upon being hacked; permissioned ledgers with single points of failure are vulnerable should anything happen to that single point, and poorly tested smart contracts can create bad consequences for participants – even if no harm was originally intended, and especially if the blockchain network does not have adequate controls in place.

C. Interledger: implementing transactions across blockchains

24. Today, many different blockchains exist and, in the future, there will be even more. Already, a supply chain transaction, from beginning to end, could involve writing or reading data from multiple blockchains. In addition, it is easy to foresee an increasing need for the exchange of information and the implementation of transactions across blockchains (i.e. interledger).

25. As mentioned earlier, blockchains can reference data outside of that blockchain. This includes data in other blockchains as well as non-blockchain systems. There are two broad categories of external data references that can occur in a blockchain system: linked data and blockchain-spanning transactions.

26. Linked data uses hashes and may also use digital identifiers and public key cryptography (as long as it is used consistently across the blockchain and whichever system the linked data is stored on). This implies that the more standardized the use of public key cryptography, the easier and less expensive it will be to link data – and the same can be said for the semantics defining the data.

27. Extrinsic blockchain references (also known as anchors) can be used to prove the existence or unchanged nature of the data pointed to. This is different from a hyperlink or Uniform Resource Locator (URL) on the Internet where the information at an address may change depending on the time it is accessed. For example, if you click on a link on a television news website, which changes on a regular basis as it is updated, what you find tomorrow may be different than what you find today. With a blockchain anchor data link, the information in the blockchain is a guarantee (proof of existence) that the data being pointed to has not been changed.

28. As well as linking data between two blockchain systems (cross-chain references) and pointing to data that may be used by a smart contract (for example a test certificate), linked data can also be used to incorporate off-chain big data into a space-constrained blockchain system. Supplementary data can either be in public/open distributed data systems such as InterPlanetary File System (IPFS – an open, content-addressable memory that uses standard internet protocols), or it may reference data in private databases that are selectively available to permissioned ledger users. With private off-chain or cross-chain references, it is possible for network operators to know that some data exists, but to have their access limited by additional controls. This can be very interesting from a privacy standpoint as it is possible to access data in order to know that, for example, someone is over 21, without giving their age, or that they live in London, without giving their address.

29. Inter-ledger (blockchain-spanning) transactions use cross-chain references and components (e.g. smart-contracts) on both blockchains that interact in a coordinated way. This is an emerging field, however there are mechanisms that already exist and are in use. These are primarily focussed on exchanging value (digital assets) between ledgers, for example Ripple interledger and the Lightning Network.