Blockchain Framework in Digital Government for the Certification of Authenticity, Timestamping and Data Property

Francesca Fallucchi
G. Marconi University
f.fallucchi@unimarconi.it

Marco Gerardi
ISPRA
marco.gerardi@isprambiente.it

Michele Petito
AGID
petito@agid.gov.it

Ernesto William De Luca
G. Marconi University
ew.deluca@unimarconi.it

Abstract

In an ever more digitized world where information and data are increasingly dematerialized, the question of how to certify intellectual property and define when a document has been created or modified without the presence of any third-party guarantor inevitably arises. This document proposes a decentralized method that, by exploiting blockchain technology and distributed peer-to-peer (P2P) networks, makes it possible to historicize information in such a way that it is not possible for a user to alter its dating, attribute ownership or modify it by impersonating the author. The data certification (document, image, film, data archive, etc.) takes place through the creation of an immutable relationship between the owner and the data. At the legal level, many countries are beginning to regulate blockchain technology to use in many areas, such as the production chain, the Internet of Things or Public Administration. In this paper, we present a solution to promote digital government and greater transparency, using a framework based on the Ethereum blockchain, a smart contract and a decentralized application.

1. Introduction

Every structure, being public and private, has a considerable amount of data and information which has to be guaranteed, according to its mandate and interest, traceability, ownership, accessibility, availability and integrity. All organizations have always needed a "super partes" organization (centralized system), such as a government or central bank, to manage mistrust and lack of relationships between counterparties. The Blockchain [1] technology, known as Distributed Ledger Technology [2][3] (DLT), helps us keep track of information, ensure data ownership and provide reliable and traceable information throughout the supply chain. This technology was originally designed to play a role primarily in the financial field [4], but in recent years it has also been exploited in other areas such as the decentralized supply chain [5], identity-based Public Key Infrastructure (PKI) [6], decentralized proof of document existence [7], decentralized Internet of Things (IOT) [8], decentralized storage [9] [10] [11].

Therefore, compared to centralized systems, the blockchain has a number of advantages:

- *Decentralization and automation of public digital archives*. Technology simplifies and streamlines data creation, management, transmission, tracking and storage flows, eliminating redundant intermediaries. Internal and external processes become faster, more efficient and transparent, with a significant reduction in operating costs and infrastructure requirements.
- *Timestamping and proof-of-existence*. Timestamps provide reliable and irrefutable proof of the existence, date of creation, origin, content, security and integrity of any existing document, contract, license, ownership, asset or event in digitized form, automatically and without human intervention. In addition, the technologies available today can also demonstrate the existence and ownership of a digitized document or asset while maintaining its confidentiality, i.e. without showing its content or owner information, with applications of relevant legal interest.
- *Tax collection and management of public funds and payments*. Digital identities, expense items and financial resources can be administered in a fully traceable and transparent manner, with no possibility of concealing or altering transactions. This makes it possible to prevent the illicit or improper use of public or private funds; to ensure that funds are spent only by certain entities or only in the manner planned, through additional applications such as smart contracts and digital multi-signature systems; to monitor more accurately the results and performance of the Public Administration.

Thanks to its characteristics, the blockchain can be used by governments to fight corruption and provide more transparency to citizens. Corruption is one of the main causes of the failure of government projects and increased public spending. For example, in public procurement, a blockchain register could record and

URI: https://hdl.handle.net/10125/70895
978-0-9981331-4-0

protect key procurement information. This would prevent tampering by a public official and allow the process to be audited. Due to the transparent nature of the blockchain, use cases can be extended to all public registers, such as real estate registers (cadaster), business registers, etc. In addition to the aspects of transparency and corruption, there are many cases of blockchain use worldwide, already implemented and in the process of implementation. An example is "The National Blockchain Roadmap" [12], a document which has been created by the Australian government and outlines a path to implement use cases that will benefit from the blockchain by 2025. One of the objectives concerns anti-counterfeiting applied to the Australian food and wine industry, which in 2017 cost over 1.68 billion Australian dollars. A study by Stanford University [13] discusses the admissibility of evidence in courts in various countries, taking China as the main reference, where blockchain and artificial intelligence have long been used to resolve disputes in virtual courts, where citizens can communicate through a screen with virtual judges based on artificial intelligence, without physically going to court.

This paper proposes a Decentralized APPlication (DAPP) [14] based on blockchain Ethereum [15][16] and InterPlanetary File System (IPFS) [10] to "certify the data" without the need of a centralized organization or system which directs and governs its management, while ensuring its ownership and inalterability. Where a company decides to publish and archive its documents using the proposed solution, for example official financial papers, communication strategies would change drastically in favor of greater transparency and a simultaneous reduction in costs. The next paragraphs will focus on the reasons that led to the choice of the technologies (Section 2), on the architecture and implementation of the DAPP (Section 3) and its implementation (Section 4). Finally, possible future developments will be presented (Section 5).

2. Related work

2.1 Blockchain technology

In recent years, decentralized cryptocurrency such as Bitcoin [17], Ethereum [15], Zcash [18] and the underlying blockchain technology are getting more and more attention. Blockchain is no longer used only in the financial sector, but finds useful applications to solve real world problems [19]. An example is Mattereum [20], a project that allows to manage legal rights of property, objects and information through the tokenisation of assets and the introduction of the concept of "automatic custodian", able to transfer property rights automatically according to terms defined by a smart contract. The same concept can therefore also be applied to intangible assets such as patents, copyrights, trademarks that once tokenized can be registered on the blockchain through smart contracts. The smart contracts, being real programs, are written with a precise logic and their execution takes place without the intermediation of a third party. For example, decentralized platforms such as MakerDAO, Compound and Dharma, use smart contracts to automate the process of paying the interest of a loan to the investor without going through a financial intermediary [21].

The tokens therefore allow to represent something unique, in this case we speak of non-fungible tokens (NFT) or not unique (for example cryptocurrency). A well-known case of NFT uses is Decentraland, a project similar to Second Life, in which users can buy non-fungible tokens (called LAND) that attest to the ownership of a property, such as land, a house, within a virtual world. Other companies like OpenSea, Known Origin and Cappasity are experimenting with NFT for digital art or digital augmented and/or virtual reality (AR/VR) experiences.

2.2 Blockchain use cases

We selected three case studies which demonstrate how the method and technologies used in our project are subject to in-depth studies and considerable funding from governments and public administrations that see these tools as useful aids for the public sector and e-government. An increasing number of governments are launching blockchain technology pilot projects and are making changes to existing laws and legislation to adapt to new technologies. Governments can leverage blockchain technology to provide information security, to optimize processes, and to integrate hyper-connected services, while building trust and accountability.

The technology behind distributed ledgers can be used for a wide range of applications, including identity management, land registration, digital currency and payments, supply chain tracking, healthcare, business registration, taxation, voting (elections and proxy), and legal entity management.

The framework illustrated in this paper could be easily applicable and declinable to such systems:

2.2.1 CS1 – United States of America. The Science and Technology (S&T) Directorate of the Department of Homeland Security (DHS) has provided a number of grants to explore innovative solutions using blockchain technology [22] for secure digital identity management [23], to improve the anti-counterfeiting capabilities of digital documentation to prevent fraud

and forgery. A Blockchain project has therefore been implemented to ensure Digital Identity as today's traditional identity systems are fragmented, insecure and unique. The Blockchain enables safer management and storage of digital identities by providing a unified, interoperable and tamper-proof infrastructure with key benefits for businesses, users and Internet of Things (IOT) management systems.

Identities stored on the blockchain are inherently more secure than entities stored on centralized servers. Using the Ethereum blockchain, in combination with distributed data storage systems such as InterPlanetary FileSystem (IPFS) or OrbitDB, existing centralized data storage systems can be disintermediated while maintaining trust and data integrity.

2.2.2 CS2- Spain. The City Council of Valls has opened to the public the Municipal Data Portal project, which publishes the datasets in the municipal web portal using blockchain and IPFS [24] technology. To achieve this goal, an application called OpenDChain, internally developed, has been developed, which replicates all the Portal's data on the IPFS peer-to-peer network, in order to guarantee its availability, and stores the hash codes on a blockchain, in order to ensure their traceability.

2.2.3 CS3 – Canada. The Government of Canada expands technological experimentation on Blockchain with the addition of IPFS technology. The National Research Council of Canada, through its Industrial Research Assistance Program (NRC IRAP), is using Bitaccess' latest product, the Blockchain Catena Suite (https://explorecatena.com), to host its blockchain explorer [25] on the IPFS network. This choice follows the success of the experiments carried out on the blockchain Ethereum in January 2018. In this project, NRC IRAP is using IPFS to host a blockchain explorer application, similar to a search engine, which will allow users to instantly search the blockchain Ethereum for grant and public contribution data.

The US case study (CS1), briefly illustrated, fully follows the possible use of Blockchain Ethereum and IPFS technology as described in this paper. In fact, the framework described in the following chapters can be used to certify the authenticity of a document in order to avoid fraud and forgery.

The Spanish design choice (CS2) also follows very well the technology used to build the framework described in this paper because the OpenDChain application, in principle, performs the same operations as the DApp developed in this project.

Finally, also in the Canadian case study (CS3) it is possible to notice that the design choices illustrated in this paper are of fundamental help for e-government in particular and for the public sector in general.

The framework developed for this project has the advantage, compared to the study cases illustrated above, of automating completely the flow of information. In fact, human intervention is reduced to the upload of the file to be saved and the authorization of the transaction. Everything else is completely managed by DAPP and smart contract, so both human error and voluntary alteration of the system are very difficult actions to perform, especially because the smart contract cannot be altered. Therefore, our framework automates the whole process of certification of authenticity, timestamping and data ownership.

2.3 Ethereum and IPFS

For the smart contract development there are a lot of performing blockchains. The most famous and used one is undoubtedly Ethereum [15] [26] [27], born after Bitcoin, with the goal of providing a Bitcoin-like system with a complete scripting language. Compared to Bitcoin, the innovation lies in a programmable blockchain. Users can create, publish and execute smart contracts on the blockchain. Recently, alternative blockchains such as Cardano [28], EOS [29], Tron [30], Ethereum Classic [31], Zilliqa [32], Lisk [33], Ontology [34], are emerging, but Ethereum is still the most solid and widespread blockchain, which is why it was adopted for this research project.

Therefore, Ethereum and the tokenization concept may be sufficient to certify ownership.

To achieve the highest level of transparency, governments could take advantage of Ethereum's "public" blockchain where any user has the ability to read data in the decentralized register and enter a transaction, if valid. In a public blockchain anyone can participate in the consent process, through which blocks are inserted into the blockchain. Since there is no central authority managing the database, the system functionality is guaranteed by a combination of economic incentives with a complex cryptographic data verification system, managed by a specific consent protocol shared by all the nodes of the network. This type of blockchain is defined as "completely decentralized".

When a greater information confidentiality is required, Ethereum can instead be used in the so-called "consortium" mode, where the blockchain nodes are managed by a pre-established group of organizations. For example, a government could create a hybrid blockchain based on Ethereum, made up of fifteen public organizations distributed throughout the country, each of which manages a node and where 10

validations are required for a transaction to be permanently written on the blockchain. The reading can be either public or private, or there may be intermediate systems where users can, through the use of specific Application Programming Interface (API), read only some information or have a restricted number of queries available. This kind of blockchain is called "partially decentralized".

In reality the blockchain is not sufficient to certify possession and ownership: in its original idea it was not designed to contain a large amount of data, as generally only the significant information of a transaction is included. Therefore, it is necessary to adopt a system that allows the storage of even large amounts of data, guaranteeing their integrity and traceability, without having to load them all on the blockchain. For this purpose, it is possible to combine a distributed file system that uses, as in this project, peer to peer technology, in order to guarantee a scalable, fault-resistant, dynamic system, optimizing performance and resource management.

Decentralized storage systems such as IPFS [10], Storj [35], Sia [9], etc. do not rely on a central service provider, but on a network of nodes where users share their disk space. These platforms provide the following advantages:

- *Increased interoperability*. The exchange of information is simplified, thanks to the adoption of standard and widespread worldwide protocols.
- *Increased security*. The centralized system represents a single breaking point, as this is inevitably being the only place where all cyber-attacks are concentrated.
- *Increased transparency*. Information access is conveyed through democratic control. For example, a provider might decide to answer a data query with partial results so as not to burden the data in a certain geographical area. This could result in economic damage to the data owner.

Among the analyzed decentralized platforms, InterPlanetary File System - IPFS [10] is the most widespread and with a more active community behind it. IPFS is an open source protocol that allows data to be stored and shared on a peer-to-peer network. This network is maintained thanks to the collaboration of community users who voluntarily decide to host content in a very similar way to BitTorrent. Unlike a web server that has all the data on central storage, in IPFS web pages are fragmented and redundant between nodes. Its routing algorithm allows the user to define the nodes to receive content and set the nodes / peers to receive files. IPFS can be used in two ways:

1. private mode, where only a limited number of nodes have the possibility to access and exchange information;
2. public mode, useful when data must be freely consulted and read by any user. In the latter case, persistence should be managed, using solutions such as IPFS and *Filecoin* clusters [36].

For this project, IPFS was therefore chosen, implemented in private mode, while the process of interaction with the distributed registry is managed by a specially developed smart contract, saved on the Ethereum blockchain. This guarantees that the result is an unalterable and completely transparent software product. To direct and manage the interaction and integration between the systems involved, was developed a DAPP that, through a web interface, has made the whole process user friendly.

3. System model

The project is based on two innovative and currently widely used distributed technologies: a peer-to-peer network for data storage (IPFS) and the Ethereum blockchain to record the unique fingerprint of the object saved in the storage system (storage).

The P2P network provides the necessary space to store information (which can be multimedia objects, documents, compressed archives, etc.) by returning a fingerprint of the object using a cryptographic algorithm that guarantees its security and inalterability. The function of the Blockchain instead is to guarantee the incorruptibility and protection of the data inserted within it from possible violations (the fingerprint of the object inserted in the P2P network), it also makes the information available in a transparent way to all those entitled to access the data. In addition to all this there is a "time validation" recognized by law that proves the existence of a specific data in electronic form at a specific date and time.

The combination of these technologies guarantee that the data has not been altered since its certification.

The interaction between user, blockchain (smart contract) and IPFS takes place thanks to a DAPP specially developed for this project. The system architecture is presented in the following figure 1:

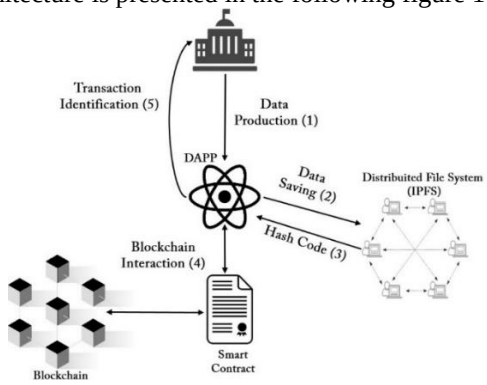

Figure. 1. System architecture

Each number of steps in Figure 1 are related to the upload process of the file to be certified, is described below:

1. The organization produces the data (file) to certify.
2. Through the DAPP, the data is saved permanently on the distributed peer-to-peer IPFS file system.
3. Once the saving is finished, the IPFS system returns the hash code of the historicized object.
4. The DAPP, exploiting the smart contract features, will permanently record the hash code on the blockchain, certifying the "data - time stamp - owner" correspondence.
5. The smart contract will return the transaction ID (Txhash) present on the blockchain to DAPP, which in turn will forward it to the data producer.

4. Case study

The following software technologies were used to implement the project:

- the decentralized platform Ethereum [15] for the creation of the private [37] blockchain and smart contract management;
- the distributed peer-to-peer IPFS - InterPlanetary Name System (IPNS) for the management and storage of data to save;
- the object-oriented, high-level Solidity programming language [38] for the implementation of the smart contract.
- the Web3.js library [39] which includes a collection of modules that implement specific functionality to interact with the Ethereum ecosystem;
- the Node.js framework [40] for server-side interaction with the blockchain and distributed filesystem;
- the JavaScript library React [41] for the frontend.
- the open source product Metamask [42] to interact with the Ethereum ecosystem (Dapp - distributed applications and Smart Contract) through a modern web browser.

In following paragraphs will be explained the measures adopted for the project implementation.

4.1. Installation and configuration of Ethereum blockchain on a local machine.

The blockchain Ethereum is a transaction-based state machine. When a block is added to the chain, the transition from one state to another takes place. Each newly generated block is linked to the previous block by a hashing cryptographic function called KECCAK-256. The main elements that make up the Ethereum system are:

- *The accounts*. The global Ethereum "shared-state" is composed of many small objects (called accounts) that are able to interact with each other through a message-passing framework. In the Ethereum platform there are two types of accounts: External Owned Accounts (EOA) and Contract Accounts (CA), both identified by a code of 40 characters (20 bytes) plus a standard prefix "0X" for a total of 42 characters. The EOA are controlled by a public/private key pair of an external person/entity, while the CA is controlled by the source code of the contract registered in the account. A message between two externally owned accounts is simply a token transfer. Instead, a message sent from an external owned account to a contract activates the contract code triggering various actions (for example: transfer tokens, write on the blockchain, mint new tokens, perform calculations, create new contracts, etc.).
- *The status of the account* (account state). Each account has an associated state consisting of a balance, a nonce (number of transactions or contracts), storageRoot and a codeHash (hash of the smart contract source code).
- *The gas and fees*. The "gas" is the unit of measure used as payment for the execution of a particular calculation on the Ethereum network.
- *The transactions and messages*. Transactions take place between accounts and vary the global status of Ethereum. All transactions, regardless of type, contain the following fields: nonce, address of recipient, gas price, gas limit, transferred ether value, sender's signature and other optional data fields.
- The blocks. All transactions are grouped in "blocks". A block chain contains a set of blocks of the same type that are linked together.

For this project has been chosen the Ethereum version written in Go (Geth) for Windows platform.

After the installation of the software product, the first account called coinbase was created which, among other activities, will mine the blockchain.

Then the blockchain was started with the geth command.

The following commands use the coinbase account and start mining the blocks in the chain:

- geth attach http://127.0.0.1:8545 (connect the console to the blockchain)
- personal.unlockAccount(eth.accounts[0], "password", 3600) (unlock coinbase account)
- miner.start() (to start mining the blockchain).

At this point, the private blockchain is active and ready to be used as the RPC HTTP server, that allows remote procedure call on port 8545, has been started.

4.2. IPFS Installation and configuration.

A prebuild package was used for the installation of IPFS on a test machine. The main implementation of IPFS is the go-ipfs program ("Go-IPFS," n.d.), written in GO language: the software contains a series of packages including an IPFS daemon server, a command line tool, an http API for node control and an HTTP gateway to provide content via a web browser. After the installation, it was necessary to initialize the repository using the shell command "ipfs init" and then "ipfs.exe daemon" to start the daemon.

The way to use IPFS is similar to the web interface. The user uploads the file to IPFS and the output is a hash string that allows the file to be retrieved. The hash string can then be compared to the Uniform Resource Locator (URL) of the web.

4.3. Smart contract implementation

The smart contract was implemented using the Solidity programming language ("Solidity Documentation," n.d.), developed specifically for the Ethereum Virtual Machine (EVM), with the help of the Remix IDE (Integrated Development Environment) available on the Ethereum.org website.

At the time of deployment, the smart contract provides the initialization of the following variables:

- *struct Record* {...}. Data structure to be inserted in the blockchain that contains a numeric identification code (id) of the object saved in IPFS (e.g. a protocol number), an alphanumeric string describing the object (e.g. photo or document), a variable (exists) used to avoid records and a timestamp (dateInsert).
- *Record records*. It represents the set of records inserted in IPFS.
- *strings[] ipfsRecords*. Array containing the hash codes inserted in IPFS.

The developed smart contract provides the following main features:

- *setRecord*. To insert in the blockchain a record containing the hash code returned by the IPFS platform plus two optional additional fields where the user can insert an identifier and a reference type of the file sent to IPFS.
- *getIpfsRecord*. Returns the data associated with the record entered by the user.

We tested the data stored in a local blockchain and the smart contract correctly returned the expected information. The smart contract was also published on the official test network of Ethereum blockchain called Rinkeby [43] at address:
https://rinkeby.etherscan.io/address/0x810C473484545A8A5BD8317de976Bfeb9F811E2f#code

4.4. DAPP Implementation

In our context, DAPP is a web application that interact easily with blockchain and IPFS. Compared to a traditional web application, a DAPP [44] should have better performance (low latency, high throughput), reasonable low transaction fee, flexible maintainability. In addition, a DAPP should not save user data locally.

For this case study was used the open source runtime environment Node.js [40], particularly suitable for the development of server-side and network applications. In addition, the web3.js [39] were used to read and write data on the Ethereum blockchain, while the react.js libraries [45] were used to manage front-end user interactions. Google Chrome was used as a web browser with the addition of the MetaMask plug-in [42] to allow users to make transactions on the blockchain Ethereum.

The application provides two main features. The first one is the loading of a file inside the IPFS distribution file system and then the saving of the unique object code (hash code returned by IPFS) inside the blockchain. The Dapp start with a form where the user can enter following information:

- file to send (mandatory);
- numeric identifier (optional);
- textual description of file type (optional).

The amount, in ether, will be taken from the wallet registered on Metamask, of which only the owner knows the private key and which will be used to sign the operation on the blockchain. After confirming the operation, a message will appear in the application to confirm the transaction, containing the following data:

1. IPFS hash code inserted on the blockchain: QmYfW4qsZdgkcwYRi9KXSntGcJi89GAU3FH7ebaNS1F47Z;
2. smart contract address: 0x810C473484545A8A5BD8317de976Bfeb9F811E2f;
3. hash code of the transaction on the blockchain: 0x5c97a9379fffc2be00a04a0a78a69e71857146ab34ad741a0f8ff0a3c28f5 b97;
4. number of the block in which the information was written: 4754108;
5. gas used for the operation: 195924.

The application then uploaded the file to IPFS and received the relative hash code of the object indicated in point 1. Subsequently, the smart contract was invoked on the Ethereum network at the address indicated in point 2. Finally, the hash code of the transaction carried out on the Blockchain was returned (point 3). The data is initially inserted on the blockchain, but is not immediately "validated" by the network, so it will be necessary to wait a few seconds

for the information to propagate throughout the network and other nodes to validate the transaction just inserted, thus making it permanent on the blockchain. After about 15 seconds will appear the data related to the block number (point 4) of the chain where the information has been saved and the cost due to the insertion of the data on the blockchain (point 5)

The second functionality provided by the application allows the search for a hash code inside the blockchain. For our test we will search the file just inserted, marked by the hash code:

QmYfW4qsZdgkcwYRi9KXSntGcJi89GAU3F H7ebaNS1F47Z

The results are returned in the form of a table with the following information:

- type of object (if valorized during insertion);
- address of the wallet who made the insertion;
- date of insertion in UNIX format;
- date converted into ISO (GMT) format;
- URL for downloading the data inserted on IPFS

Of particular importance is the presence of the address of the wallet that made the insertion on the blockchain because it guarantees and binds the owner of the wallet to the inserted object in an incontrovertible way. Moreover, if someone subsequently tried to insert the same hash code on the blockchain, the smart contract, according to the code that has been made, would block this operation because the data is already present.

4.5. Storage analysis

The results have shown how it is possible to store large amounts of data and guarantee its authenticity, timestamping and ownership thanks to the integration of a peer to peer network to historicize the data and the use of blockchain technology to save the digest. In fact, if we would use a public network such as Ethereum for example, it would be prohibitive to store even just a few kilobytes file on the blockchain because the costs would be abnormal. If we make a simple calculation, based on the current cost (October 2019) of the gas which is 8 gwei, to write 8 bytes on the blockchain would require 160,000 gwei. That is about 0.02 Ether/kb or 20 Ether per megabyte of data. Therefore, archiving a 1megabyte file would cost (at the current exchange rate: 1 ether = 160 €) about 3,200.00 €. Using instead a peer-to-peer network to save the data, storing only the digest of the object on the block chain would have a very low cost (about 0.0016 Eth = 0.25 €). The proposed solution provides significant cost savings, and avoids unnecessary redundant copies of data on every single node of the blockchain, saving storage space.

4.6 System performance

In order to calculate the performance of developed system is necessary to analyze its three main macro functionalities listed below (fig. 2):

1. file upload to DAPP (step 1)
2. file transfer on IPFS platform (step 2)
3. insertion of hash code returned by IPFS on the Ethereum blockchain (step 3)

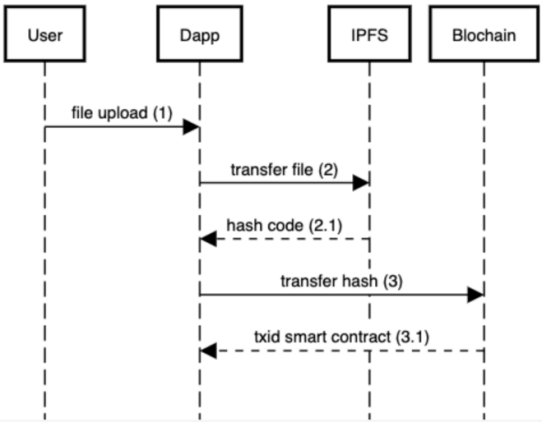

Figure. 2. System sequence diagram

In point 1 the transfer time is mostly related to the bottleneck represented by the information transmission system between the user and the server on which the DAPP is installed.

Regarding point 2, considering that there is an IPFS local node, the loading time of the file and the generation of the relative hash code takes very few seconds and is therefore negligible. The propagation time of the data just inserted on the IPFS nodes outside of the local network infrastructure took on average between 1 to 5 minutes before information became available at: https://ipfs.io/ipfs/<hashcode>.

In point 3, from the user's point of view, the performance of the blockchain is relative to the speed with which the requested service is obtained. A transaction is represented by a data entry on the blockchain, invoking a smart contract function. Therefore, the DApp prepares a transaction, sends it to the network and expects the response processed. This time frame represents the time within which the client will receive a reliable confirmation that the transaction has been executed correctly and confirmed by the network. The real bottleneck of these steps is the congestion of the blockchain network and the expense that can be incurred. The blockchain Ethereum, on average, mines a block every 15 seconds, so on the Rinkeby test network the average waiting time varied in the range [15, 20] seconds before the data is written on the blockchain. If a private Ethereum network or a network belonging to a consortium (permissioned

blockchain) is used, the time to execute the transaction would be very similar to that found on the test network. If a public network is used to save the hash code, the price to pay in terms of "gas" becomes fundamental. On sites such as Gas Station [46] and Etherscan [47] are available market quotes which indicate how quickly (on average) the miners enter transactions into the blockchain. By paying a price close to the maximum market price there is a chance that the transaction will be executed in about 30 seconds. In conclusion, it is possible to say that the performance of the system largely depends on the speed of the Internet connection provided by the ISP and the blockchain type intended to be used.

4.7 Security analysis

The proposed system pays great attention regarding the data security; in fact, the main concern is to ensure security against potential threats that could compromise authenticity, timestamping and ownership of the data.

For this purpose, decentralized and distributed technologies (blockchain and IPFS) have been used which provide the following security features:

High availability and reliability

The blockchain Ethereum replicates the data on all the nodes of the network geographically located throughout the globe. There is no single point of failure and any damage to a peer does not affect the overall functioning of the system, unlike the way it is accessed on centralized systems where a possible computer attack or blackout of the data center can interrupt the service of the entire system. This redundancy gives the entire network high availability and reliability as well as being highly resistant to failures or cyber-attacks. As far as the IPFS network is concerned, similar reasoning is possible even if the data is replicated only on some nodes of the peer-to-peer network, unlike what happens on the blockchain.

These replicas are organized in a structured way taking into account the needs of availability, reliability, load balancing, geographical location and use of the information. Moreover, data retrieval is guaranteed within a limited number of hops;

High information integrity

All transactions on the blockchain, the IPFS network, and its connections are encrypted using the most secure and modern security protocols to date, such as the Transport Layer Security (TLS) [48] and Hyper Text Transfer Protocol Secure (HTTPS) for the connection between decentralized systems and the DApp, thus giving the system high integrity.

Unchangeable data (Data Integrity)

When data have been recorded in the blockchain, it is extremely difficult (if not impossible) to modify them as this would involve a huge expenditure of computing power. For the IPFS network, the assets are identified by a globally unique identifiers (GUIDs) derived from a secure hash of the asset itself. The use of a hash code makes the resource "self-certifying", anyone can verify its validity and this protects it from any tampering, as any modification, even if minimal, changes the fingerprint. Therefore, P2P storage systems are intrinsically very suitable for storing "unchangeable" objects (such as music files, videos, images, legal documents, etc.).

Data Confidentiality

All communications between Dapp and other systems are encrypted using the TLS protocol which makes it difficult for a malicious user to tamper with communication data. Only the authorized host has access to the encrypted data and the communication channel that has been established. Smart contracts use advanced data encryption systems that are essential to ensure the highest standards of security, the resulting level of protection makes them safe to use even in mission critical processes.In addition, because smart contracts are accurate and secure and their level of efficiency is high, this represents a benefit for securing transactions.

DAPP

The Dapp acts as a control element that regulates data transfer between the IPFS network and the blockchain. It does not directly manage, manipulate or store the data it receives and therefore the risk of compromise is relatively low (especially if the DApp is installed on a corporate server that complies with business security standards and policies).

In addition, for interaction with the blockchain is required the private key to sign the transaction, but this information is in no way present on the server because it is only client-side information. Some potential attacks against the proposed system are:

51% attack risk. If a network node controls more than 50% of the computational power of the entire network [16], it would be able to modify the distributed registry, for example by eliminating some transactions or spending twice the same cryptocurrency. Although it is very difficult to obtain more than 50% of the network's computing power, this is theoretically possible and large corporate criminals could combine their capabilities to operate in that direction.

Bug Software. The developed software needs a careful internal and external audit to minimize any security holes. Furthermore, companies that embark on digitization processes must be aware of the risks arising from the use of information technology and put in place all the necessary security policies to avoid data and security leakages.

Human error. A potential risk factor is represented, by the human being who, willingly or unwillingly, puts the systems integrity at risk by carrying out wrong operations. The Dapp tries to limit as much as possible the interaction with the human being but unfortunately, it cannot guarantee, for example, that the user's private keys will not be lost or stolen.

5. Conclusion and future work

The increasing use of text, audio and video documents in digital format and the possibility for them to be copied and modified creates a new problem: how can we certify when a document has been created or modified by indicating without any doubt its author and guaranteeing its integrity?

The certification method has to meet the following criteria [49]:

- guarantee the authenticity of the data, i.e. reassure that no one has modified even a single bit of information since its creation;
- mark a datum irrefutably with the time stamp to ensure that it existed in electronic form on a specific date and time;
- allow to trace back to the author of the product without a shadow of a doubt and without any fraudulent attribution of authorship.

The proposed solution foresees the use of IPFS network to save the digital object to be certified in order to make it freely accessible in case of a public network, or to limit its use to authorized users only in case of confidential information, through the use of a hybrid or private network. Once saved the data within the distributed network, the network returns the object's fingerprint called digest. The digest of a document is realized through a unidirectional hash function that, in computer science, is used to encode a sequence of variable length data (documents, photos, archives, films ...) in a fixed length string. At this point it will be possible to save inside the blockchain the fingerprint of the object to be certified instead of the original document. The implemented solution guarantees the following advantages:

- Openness and transparency as there is no centralized body that holds a form of power and control over the entire system.
- Greater security compared to centralized systems as a decentralized system is very secure and difficult to attack both from outside and inside the network.
- Enhanced resistance to corruption as public officials would have no chance to alter data within the blockchain
- Speed and efficiency because the data are easily accessible by those entitled without the need for intermediaries that necessarily slow down the information and decision-making processes.

There are many benefits of using blockchain in the public administration where different use cases range from open data portals, publication of public registers to trial evidence registration. Governments can finally guarantee full transparency in their operations, due to the sharing of public registers by adopting a blockchain with a fully permission-less configuration such as the Ethereum blockchain.

The solution proposed in this paper could be a valuable aid to be able to "certify the data" without the need of a centralized organization or system which directs and governs its management, while ensuring its ownership and inalterability. The most relevant issue on the Ethereum blockchain is the gas price [50]. Developments are in progress and will most probably solve the problem. One of the proposed solutions is represented by the improvement proposal (EIP) 1559 [51], while on another front is working on Ethereum 2.0 that will use proof-of-stake consensus and follow a model based on the possession of tokens, therefore less expensive, and more eco-friendly. Future work will cover an in-depth analysis of other emerging blockchains that can solve the problem of the very high cost of GAS and perform faster transactions in the Ethereum network. For example, the EOS and Tron blockchains offer services with reduced costs and good performance. Therefore, software re-engineering (DAPP and Smartcontract) will be required in order to adapt it to the new ecosystem.

References

[1] Antonopoulos A. M., Mastering Bitcoin: Programming the open blockchain, O'Reilly Media Inc., 2018

[2] Bashir I., Mastering Blockchain,2nd Ed., Packt, 2018

[3] CEN-CENELEC Focus Group on Blockchain and Distributed Ledger Technologies (FG-BDLT) White Paper Subgroup: N 001, 2018.

[4] Blockchain for Financial Services. (n.d.), retrieved March 1, 2020, retrieved May 23, 2020, from https://www.ibm.com/blockchain/financial-services

[5] Blockchain for Supply Chain. 2019, https://www.ibm.com/downloads/cas/D2L1BJVA

[6] Fromknecht, C., Velicanu D., A Decentralized Public Key Infrastructure with Identity Retention, Cryptology EPrint Archive, 2014
[7] Proof of Existence. (n.d.), retrieved April 21, 2020, from https://proofofexistence.com
[8] A Decentralized Network for Internet of Things (n.d.), retrieved April 20, 2020, from https://iotex.io
[9] Vorick D., Champine L., Sia: Simple Decentralized Storage, 2014. https://sia.tech/sia.pdf
[10] Benet J., IPFS - Content Addressed, Versioned, P2P File System, 2014, retrieved May 23, 2020, from: https://arxiv.org/abs/1407.3561
[11] Zyskind G., Nathan O., Pentland A. S., Decentralizing privacy: Using blockchain to protect personal data. Security and Privacy Workshops, SPW 2015.
[12] Australian Government, THE NATIONAL BLOCKCHAIN ROADMAP: Commonwealth of Australia, 2020, retrieved Sept 27, 2020, from: https://www.industry.gov.au/sites/default/files/2020-02/national-blockchain-roadmap.pdf
[13] Polydor S., Blockchain Evidence in Court Proceedings in China, 2019, retrieved Sept 27, 2020, from https://stanford-jblp.pubpub.org/pub/blockchain-evidence-courts-china/release/1
[14] Raval S., Decentralized Applications: Harnessing Bitcoin's Blockchain Technology (1st ed.), O'Reilly Media, Inc.,2016
[15] Wood G., Ethereum: a secure decentralized generalized transaction ledger, Ethereum Yellow Paper, 2014. https://doi.org/10.1017/CBO9781107415324.004
[16] Buterin V., Ethereum White Paper. 2014
[17] Nakamoto. S., Bitcoin: A Peer-to-Peer Electronic Cash System, 2008, retrieved from https://bitcoin.org/bitcoin.pdf
[18] Hopwood D., Bowe S., Hornby T., N. W. (2020). Zcash protocol specification, retrieved from https://raw.githubusercontent.com/zcash/zips/master/protocol/protocol.pdf
[19] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H., An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends, Proceedings, IEEE, BigData Congress 2017. https://doi.org/10.1109/BigDataCongress.2017.85
[20] Mattereum, Mattereum White Paper, 2020, retrieved April 30, 2020, retrieved from https://mattereum.com/wp-content/uploads/2020/02/mattereum-summary_white_paper.pdf
[21] Delphi Digital, Decentralized Finance (DeFi): Thematic Insights, 2019, retrieved from: https://www.delphidigital.io/defi
[22] Homeland Security, DHS S&T Announces New Collaborative Blockchain Innovation Solution, 2018, retrieved Sept 27, 2020, from: https://www.dhs.gov/science-and-technology/news/2018/12/04/news-release-st-seeks-collaborative-blockchain-innovations
[23] Consensys, Blockchain in Digital Identity, 2019, retrieved Sept 27, 2020, from: https://consensys.net/blockchain-use-cases/digital-identity/
[24] Valls City Council Open Data Portal, retrieved Sept 27, 2020, from: https://dadesobertes.valls.cat/en/about
[25] NRC-IRAP - Blockchain publishing prototype, 2019, retrieved Sept 27, 2020, from: https://nrc-cnrc.explorecatena.com
[26] Ethereum Homestead Documentation. (n.d.). retrieved April 3, 2020, from https://readthedocs.org/projects/ethereum-homestead
[27] Ethereum Blockchain App Platform. (n.d.), retrieved May 3, 2020, from https://www.ethereum.org
[28] Cardano. (n.d.), retrieved March 2, 2020, from https://www.cardano.org/
[29] EOS, retrieved March 22, 2020, from https://eos.io/
[30] Tron, retrieved 23/4/2020, from https://tron.network/
[31] Ethereum Classic (n.d.). https://ethereumclassic.org/
[32] Zilliqa (n.d.), retrieved April 3, 2020, from https://www.zilliqa.com/
[33] Lisk. (n.d.), Retrieved May 3, 2020, from https://lisk.io/
[34] Ontology (n.d.). https://ont.io/
[35] Wilkinson S., Boshevski T., Brandoff J., Prestwich J., Hall G., Gerbes P., Pollard C, Storj A Peer-to-Peer Cloud Storage Network, Storj.Io, 2016
[36] Benet J., Greco N., Filecoin: A Decentralized Storage Network, Protocol Labs, 2018
[37] Buterin V., On Public and Private Blockchains. https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/, 2015
[38] Dannen C., Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners, Apress, 2017.
[39] Web3.js - Ethereum JavaScript API, retrieved April 3, 2020, from https://github.com/ethereum/web3.js/
[40] Young A., Meck B., Cantelon M., Node.js in Action, Manning, 2018
[41] Sriparasa S. S., JavaScript and JSON Essentials, Packt Publishing, 2013.
[42] Metamask. (n.d.). http://www.metamask.io
[43] Rinkeby, https://www.rinkeby.io/#stats
[44] Cai et al., Decentralized Applications: The Blockchain-Empowered Software System, IEEE, 2018, retrieved 23/5/2020 from https://arxiv.org/pdf/1810.05365.pdf
[45] React, A JavaScript library. retrieved April 4, 2020, from https://www.reactjs.org/
[46] Eth Gas Station, https://ethgasstation.info/
[47] Gas Tracker, https://etherscan.io/gastracker
[48] Internet Engineering Task Force (IETF), RFC7919, 2016. https://tools.ietf.org/html/rfc7919
[49] Report Finale, Internet Governance Forum (IGF) – Nazioni Unite, Italia, 2018. https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2019/02/01/igf-italia-online-il-report-delledizione-2018
[50] Sassano A., Creating a Symbiotic Relationship, 2020, retrieved Sept 27, 2020, from: https://thedailygwei.substack.com/p/creating-a-symbiotic-relationship
[51] Ethereum improvement proposal (EIP) 1559, retrieved Sept 27, 2020, from: https://github.com/ethereum/EIPs/blob/master/EIPS/eip-1559.md