NFB: A Protocol for Notarizing Files over the Blockchain

Haikel Magrahi, Nouha Omrane, Olivier Senot, Rakia Jaziri

NFB: A Protocol for Notarizing Files over the Blockchain

Haikel Magrahi
Paris VIII University
IRT SystemX, Paris-Saclay, France
Student and Data Engineer
Email: megrhi.haikal@gmail.com

Nouha Omrane
Docapost DPS, France
R&D Expert
Email: nouha.omrane@docapost.fr

Olivier Senot
Docapost DPS, France
Director of new services
Email: olivier.senot@docapost.fr

Rakia Jaziri
LIASD, Paris VIII University
Associate professor
Email: rjaziri@ai.univ-paris8.fr

***Abstract*—Blockchains such as Bitcoin and Ethereum and their respective P2P networks have seen significant adoption in many sectors in the past few years. All these technologies that follow the Blockchain pattern shown that its possible to rebuilt any transactional system with better performance without relying on any trusted parties to manage transactions between peers. This idea made many companies invest millions to understand this technology and to find a way to immigrate from the centralization to the decentralization.**

These solutions need to store large amounts of data in a secure and confidential way. Many storage systems have emerged to the blockchain technology, by storing the data in a distributed storage systems (Cloud, FileSystems, etc). But none of these tools proposed are able to manage document's life cycle nor archive documents based on regularity compliance.

In this paper, we describe our protocol called NFB (it stands for Notarizing Files over the Blockchain). This protocol ensures the communication between two systems: a permissive Blockchain and a secured centralized Document Management System. The described method is used to allow users to archive, control, analyze and validate their transactions in a system that offers confidentiality, security and distribution features.

Keywords*—*Blockchain, Quorum, Archiving documents, Regularity compliance, Protocol, DMS-ECM.

I. Introduction

At first, Blockchain was dominating financial sectors slowly, by offering trading, exchanging, supply-chain securely and efficiently without needing any central point of control. Then, cryptocurrency blockchains and their respective P2P networks start to be useful beyond exchanging money which made various industries think of exchanging objects not just cryptocurrencies and build decentralized application introducing the Blockchain 3.0.

Nowadays, the Blockchain technology is used in many sectors such as Media, Health-care, Education, etc. More precisely, this technology is used along with other applications like File Storage or Document Time-stamping.

In this paper, we are interested in Data archiving field. Recently, many solutions have been released such as Storj, IPFS [7], Blockstack [9] [9], etc. Despite their decentralized approach, they still require a set of rules to ensure the long term storage of documents and their integrity. Besides, The decentralized model poses some challenges when there is a need to verify user identity or to ensure document access confidentiality.

A Data Management tool is therefore important for the implementation of sustainable operations related to the management of the document's life cycle in order to archive documents with a legal compliance. In this paper, we present our protocol NFB which relies on a centralized archiving solution called OKORO which is certified ISO 14641-1.

II. Blockchain technology

Since the creation of the Bitcoin Blockchain [2] in 2009, many types of blockchain has been developed. Some Blockchains are designed to be completely public (open and permission-less) and they expose the ledger to all participants, such Bitcoin and Ethereum. Other blockchains are permissioned and all its information are encrypted in a distributed ledger, such as Quorum and Hyperledger [3].

Blockchains 2.0 expose a collection of business rules called smart contracts which are deployed in the network, distributed and validated by all the peers in the network. A smart contract can be highly serviceable in automating business processes in a trusted way by allowing all stakeholders to process and validate contractual rules as a group which increases confidence and security. In this section, we compared the most famous Blockchain technologies.

1) Ethereum: Ethereum is the second most popular public Blockchain for exchanging assets and payment that is called Blockchain 2.0. Ethereum supports Smart contracts with three different languages (e.g. Solidity, snake, etc). Also, its a protocol for building decentralized applications. Each transaction is paid with cryptocurrency called ETH, and has a Gas (which used to determine how much ETH you need to pay to send a transaction). To validate the transactions, Ethereum supports Proof of Work: a consensus-based on Miners (any participants can become a miner). For private networks, to

test your decentralized application, Ethereum supports many consensuses like POW, POS, BFT, etc, and an SDK named Web3 to manage the Blockchain, send transactions, etc.

2) Multichain: Multichain [5] is one of the first permissioned Blockchain for financial sector by providing privacy and access control list. It offers a private Blockchain that can be used as a package and its easy to deploy either within or between organizations and for any different operating system from Linux to Mac OS. Multichain doesnt support smart contract for performance reasons.But, on the other hand, it offers a set of assets (like sharing transactions, sending transactions, etc) to manipulate them. For the transactions validation, Multichain uses the PBFT (Practical Byzantine [1] Fault Tolerance) what makes it more performed than using Miners.

3) Quorum: Quorum [8] is an Ethereum fork and a private Blockchain developed by JP Morgan. They modified Ethereum by adding permissions, private transactions and they changed the consensus from the POW to a voting system (implemented in a smart contract), which boosted transactions per second rate to the limit.

4) Hyperledger Fabric: Hyperledger is a permissioned shared ledger that offers high degrees of confidentiality, resiliency, flexibility and scalability hosted by Linux Foundation. Hyperledger is developed in a modular architecture which makes its modules reusable and pluggable. Hyperledger offers a certified authority server for enrollment and revoking users from the Blockchain (by delivering ECert and TCert). Hyperledger supports smart contract named chaincodes with two languages (Java and Go) and channels which are a sub-blockchains. It doesnt support cryptocurrencies nor gas mechanisms. Hyperledger uses PBFT consensus by separating the validation into different roles: Commiter, Endorser and Orderer. Also, Hyperledgers ledger could be CouchDB or LevelDB which facilitate querying the Blockchain.

These Blockchain technologies don't support file storage. In fact, many storage decentralized tools were proposed to store documents that are exchanged between peers in the Blockchain network. In the next section, we detail the most famous storage tools in a P2P network Blockchain.

III. Storage systems in the Blockchain

As described in the previous section, the Blockchain is a peer-to-peer distributed ledger shared among all the peers in a network. This technology, first has gained the trust of the financial sector by offering exchanging, trading, managing assets efficiently and securely. In the recent years, it was introduced in many other sectors such as healthcare, insurance and automobile by offering cost-efficient without needing a central point of control.

The distributed ledger contains all validated transactions made between peers in the P2P network. But the data exchanged between users (e.g. files) are not stored in the ledger. For that reason, numerous storage solutions and protocols were created during the revolution of this pattern by offering a more secure and decentralized storage solution without relying on the central control point.

Blockchain-based decentralized storage is cheaper, more secure, faster, censorship resistant, and more distributed than existing cloud storage solutions. In this section, we mention some of the popular solutions that follow the Blockchain's pattern.

A. StorJ

Storj [6] is a Blockchain-based cloud storage. It offers client-side encryption which increases the security. Also, it offers the possibility of transferring and sharing data without depending on a central point of control. More precisely, Storj secures documents by encrypting and splitting them into shards (or partitions) selected by the user. Every shard is a peer in the network. This mechanism is similar to BitTorrent. Unlike traditional Cloud storage providers, you can access your documents securely anytime as long as you have the key.

B. IPFS

IPFS [7] is a Blockchain based File-system that is distributed in all peers in the network. IPFS combines Git, BitTorrent and the Self-Certified File-systems. It offers exchanging files like Bit-Torrent but uses versionning like Git.

IPFS affects a cryptographic Hash to each file. Then, it removes duplicated files and gives them a version by using git mechanism (version history). Next, each network node stores a copy and some indexing information for search optimization.

C. Blockstack

Blockstack [9] is a new decentralized internet where users keep their data encrypted before it will back up it in the cloud. This removes the need for blind trust in 3rd parties and makes it easier to keep your data safe. Applications run locally. In fact, they are loaded via a secured domain name system and exist on your device.

As we discussed in the last sections, all these solutions are for storing documents, by replicating and distributing them across the network. Applying this techniaues increase the high availability and performance. But in the other hand, it makes documents less secured and exposed in different nodes in the network.

Meanwhile, archiving is not just storing and saving data in a storage solution. Documents must pass by a set of activities, contracts that will determine its life cycle and these data must be secured in known places by clients. Besides, most of time, archiving tools are legally compliant which are not the case in these blockchain storage solutions.

In the next section, we describe our protocol NFB that is built on top of the Quorum Blockchain [**?**] and uses OKORO as a centralized DMS-ECM.

IV. NFB Protocol

Notarize File over the Blockchain Protocol (NFB) guarantees the communication between two different ecosystems: the Blockchain and a centralized archiving solution (DMS-ECM). It offers major services such as document's archiving, document's retrieving and document's proof existence thanks to the use of a DMS-ECM for archiving documents and

the blockchain to trace transactions related to the archived documents.

NFB is a protocol that can be integrated easily with any given centralized DMS-ECM solution (such as OKORO, Alfresco, SAP DMS-ECM, etc) and with any given Blockchain based solution (e.g. Hyperledger, Ethereum, Quorum, etc).

A. Archiving documents

This service accepts an encoded documents associated to a collection of meta-data that represent a set of information (e.g. title, date, etc). Documents and their meta-data are stored and indexed in the DMS-ECM. The process controls and validates document's format and its meta-data in the DMS-ECM side. Then, NFB traces all the transactions related to these actions in the Blockchain. Several information are traced in the distributed ledger. Some information are related to the transfert process (e.g. public address of the sender, transaction's timestamp, etc). Other information correspond to the Data Archiving System such as document's ID, document's hash, file's name, etc.

B. Retrieving documents

Besides archiving documents, NFB offers the possibility of retrieving documents in order to enable searching and downloading documents physically. Only users that have the right to search and access to a set of documents can operate these actions. The retrieving process accepts one or several keyword that correspond to user's queries. Keywords are related to either document's meta-data such as date, document's hash,or information that are traced in the Blockchain such as document's ID or the transaction's label.

The result corresponds to one or several documents according to the query's result tuned in the DMS-ECM side. Users can download the selected documents, if they have to right to operate this kind of action. At the same time, NFB traces these actions in the Blockchain. Transactions, that are traced in the distributed ledger, contain several information such as Document's ID, retrievel's action timestamp, etc.

C. Document proof existence

The protocol NFB enables auditing documents publically in order to prove their existence through time. If one person wants to verify a document's integrity and existence, the protocol offers the possibility to compute document's hash and verify document's existence thanks to the correspondence between information that are permanently traced in the Blockchain (e.g. document's hash) and those related to a document archived in OKORO.

More precisely, NFB offers a real time streaming process for transactions and blocks that are validated in the Blockchain for those who want to monitor transactions in the Blockchain. Users can easily search for transactions, related to a document's proof existence, by quering the Blockchain using keywords like document's hash, timestamps or by transaction's hash.

Users can select a transaction from the resulting output list. Then, once selecting a transaction, users can check a document's information that were traced in the Blockchain (e.g. document's hash) and other indexed in OKORO (e.g. document title). This proves that a document was well archived. Otherwise, NFB matches zero transactions for users queries.

In the next section, we go more deeper into our NFB protocol by exploring its architecture and the modules that make it easy to maintain and to extend.

V. NFB ARCHITECTURE

The implementation of our protocol NFB relies on the Micro services architecture standard. The figure 1 describes our prototype in a high layered level.

This prototype contains three major modules. The first module defines all Blockchains dependencies such as the SDK, smart contracts' code source, etc. Also, it implements all the functionalities that enable the writing and reading in/from the Blockchain such as deploying smart contracts, sending transactions between peers and even retrieving blocks.

The second module, named "OKORO layer", includes all DMS-ECMs dependencies. It defines a set of web services to manage documents through an archiving centralized solution (named OKORO). Such defined functionnalies are opening a stream, injecting documents and retrieving documents. This layer can contain web services endpoints or even other methods like Remote method invocation (RMI).

The third module plays the role of a gateway-like. It synchronizes and orchestrates the tasks between the Blockchain layer and the DMS layer. Besides, it manages authentication and authorization, returns web ui for users to test the protocol and defines session management.

Using this modular application made our protocol easy to reuse and enable code source's update easily by offering a set of interfaces that you need to re-implement in order to change the Blockchain technology or the used DMS-ECM.

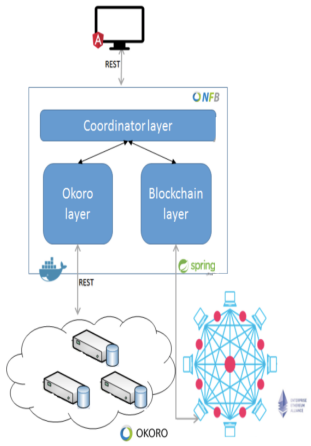

Fig. 1. NFB architecture

A. Proof of Archivability

As we explained in the NFB Protocol section, this service archives documents in a centralized trusted archiving solution and traces these actions' request and response in the Blockchain

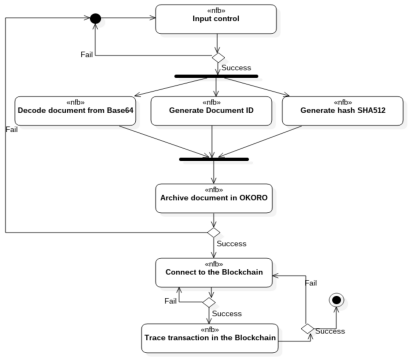

Fig. 2. Archivability services activity diagram

B. Proof of Retrievability

As we explained in the NFB Protocol section, this service offers the possibility of searching and downloading documents from a centralized trusted archiving solution and also tracing request's and the response's actions in the Blockchain 3.

(1) Coordinator layer loads the keyword introduced by a user and sends a search request that contains a set of keywords to the DMS layer.
(2) The coorfinator layer establishs a connection to the centralized archiving solution (OKORO) and starts retrieving the related documents by processing a set of web services. At the same time, the coordinator layer (3) establishs a connection based on the user's address by sending a request to the Blockchain layer.
The Blockchain layer (4) uses a specific SDK to connect to the Blockchain or a set of custom web services to trace that action with some other information in the Blockchain. When the downloading process terminates(7), The DMS layer will format and forward the response to the coordinator layer. If a user got the response, it will reprocess the same steps (3,4,5,6) to trace the response action into the Blockchain.

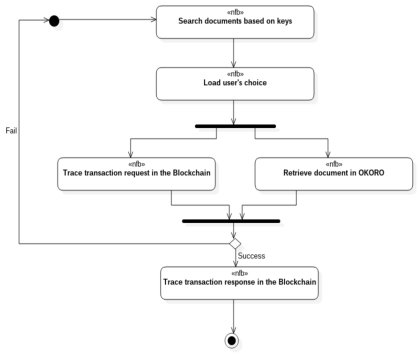

Fig. 3. Retrieving services activity diagram

C. Proof of Existence

As we explained in the NFB Protocol section, this is the most important service in our protocol. It proves a document's existence and integrity by validating the existence of a documents transaction in the distributed ledger. It means that this service provides a transaction related to a document that already exists in OKORO. In other words, POE selects all meta-data related to a document within the transaction and search it in OKORO. When the user wants to prove the existence of a document, NFB traces that action in the Blockchain too. This service includes many search possibilities such as users address, documents hash, documents name, action's label or even by loading a document (to calculate it's hash).

The figure 4 shows the activity diagram of the document proof existence service.

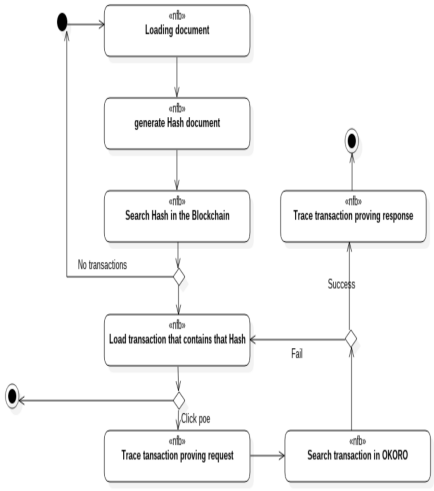

Fig. 4. Document proof of existence's activity diagram

In the next section, we detail the technologies used to implement the protocol NFB and we describe some screenshots of the frontend application.

VI. Experiment

In this section, we explain all the technologies that were chosen in the implementation of NFB Protocol.

A. DMS-ECM Solution - OKORO

OKORO is a document management solution (DMS-ECM), designed for ETIs and large accounts, and was introduced by Docapost DPS as a document archiving solution with regulatory compliance. This solution offers many services such as improving the operational efficiency and productivity of the company (ease of use with an ergonomic interface containing all the documents of the company, saving time with the instant search of documents). It enables access to all documents immediately wherever you are (sharing and easy transfering of information in the company) thanks to the SaaS mode.

All the proposed web services are secured (full tracability, optimized backup of sensitive documents, etc). Besides, OKORO was certified within many labels such as ISO 9001, FNTC-TA, FNTC-CFE 2013, ISO 14641-1 and ISO 27001.

B. Blockchain technology: Quorum

Firstly our choice was to use the Hyperledger Blockchain as the perfect choice in our POC by offering a real database as a ledger which ameliorates storage capacity and queries. But the problem was that Hyperledger was underdevelopment and there wasnt a stable version. Thats why we opt to use the Quorum Blockchain as a Blockchain P2P network to trace transactions related to documents archived in OKORO.

C. Other technologies

Technologies used in our PoC were chosen by considering the richest community, the maturity of the technology and its performance.

1) Java Spring: An open source project for building Java web application. Spring hides all Java EE and web complexity. Its famous thanks to its inverse of control design pattern which makes developers manipulate components and inject them when needed. This decreases dependencies and increases the ease of re usability and testing.

Also, Spring offers sub-projects to facilitate many tasks like integration by offering Spring integration project, databases mapping by Spring data, etc.

2) JavaScript: JS is a weakly, dynamic typed programming language and presents the core of all the technologies World Wide Wed along with HTML and CSS. In the old era, JS was used only for animations, input control and to offers interactive pages. Until some frameworks have built on it like AngularJS and ExpressJS which take JS to another level by adding architectures and asynchronous programming.

3) Docker: Is a software technology affording containers, developed by the company Docker.Inc. Docker provides an extra layer of abstraction of operating-system-level virtualisation on Windows and Linux. It isolates your application deployed in the container from your operating system which facilitates deployment.

D. Results

NFB Protocol can be used in different sectors such as automobile, insurance for documents archiving purpose. We deployed and tested this solution in a VM Openstack using Ubuntu server with 4vCPU, 8GB RAM, and 30GB storage.

Our solution offers both web-ui (user interaction) and web services endpoints to be integrated as services with other tools. We Test our protocol as a module integrated in a use-case related to a solution proposed by PSA group in the automobile sector. In this use-case, users can archive their documents using NFB and verify document's existence publically.

This shows an easy integration of our PoC with the existing solution and good results by guaranteeing archiving and retrieving documents and tracing the corresponding transactions in the Blockchain. NFB protocol traces all the archiving/retrieving actions in the Blockchain and archives the physical documents in OKORO. -

APPENDIX A
PROOF OF ARCHIVABILITY

The user introduces the document's meta-data and then selects the document(s) to archive. This ui is related to the DMS-ECM OKORO. Figure 5 shows a simple ui used to demonstrate our POC to the clients.

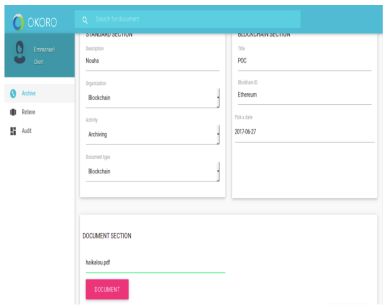

Fig. 5. NFB Web-ui archiving documents

APPENDIX B
PROOF OF RETRIEVABILITY

The user introduced some keywords for searching, such as a date, document's hash, etc. Like we mentioned earlier, NFB searchs for the documents that match the keyword and displays all the documents. Then, the user selects the document to download it. Figure 6 shows a simple ui used to demonstrate our POC to the clients.

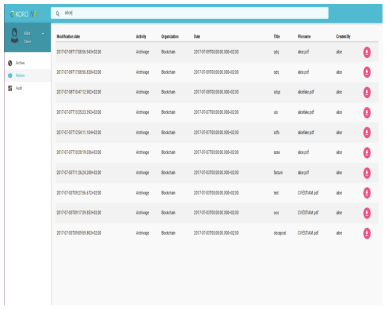

Fig. 6. NFB Web-ui retrieving documents

APPENDIX C
PROOF OF EXISTENCE

The user searchs for existing transactions by any keyword (e.g. transaction's hash) (Figure 7). After selecting one transaction from the ouput result, user can consult the information related to that transaction (Figure 8). Then, the user can click on the button POE to prove if that document related to that transaction is well archived in OKORO or not (Figure 9). The response is a set of informations that exist in OKORO and that are related to the given document.

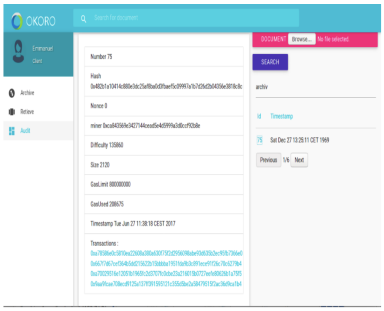

Fig. 7. NFB Web-ui 1 proving documents' existence

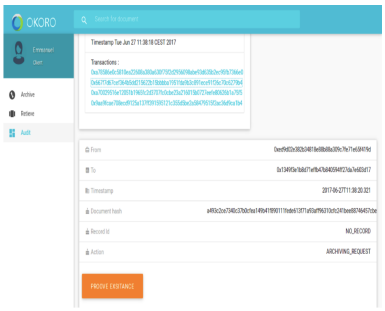

Fig. 8. NFB Web-ui 2 proving documents' existence

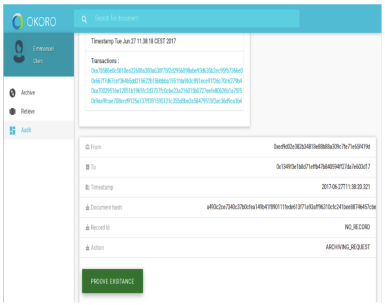

Fig. 9. NFB Web-ui 3 proving documents' existence

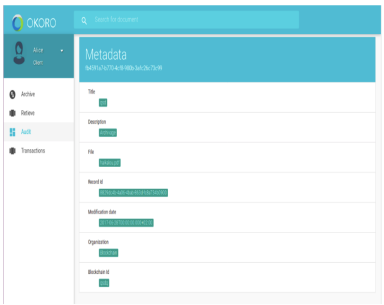

Fig. 10. NFB Web-ui 4 proving documents' existence

Conclusion

NFB is a protocol that guarantees a communication between a centralized archiving solution named OKORO and the Blockchain in order to notarize files over the Blockchain.

NFB offers three major services. The first one is archiving documents in a centralized secured solution and tracing transactions that correspond to this action in the Blockchain. The second service is retrieving documents by offering the possibility to search document(s) and download them from a centralized secured solution (OKORO). Besides, the protocol traces all actions in the Blockchain thanks to transactions.

Finally, our protocol allows document's proof existence. This service is a fundamental functionality that our protocol offers to users. It proves that a document exists (or not) by returning its metadata indexed in the secured centralized solution based on information traced in the Blockchain, during a period. NFB was integrated in the PSA Peugeot use-case to archive invoices and to trace users' actions in the Blockchain.

References

[1] Aublin, P., Mokhtar, S.B., Quma, V. *(2013). RBFT: Redundant Byzantine Fault Tolerance.* Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on, pp.297-306. doi:10.1109/ICDCS.2013.53 or http://dx.doi.org/10.1109/ICDCS.2013.53

[2] Satoshi Nakamoto, 2008 *Bitcoin: A Peer-to-Peer Electronic Cash System*

[3] Linux foundation, *Hyperledger: Whitepaper* http://www.the-blockchain.com/docs/Hyperledger-Whitepaper.pdf

[4] Ethereum, *Ethereum: Whitepaper* https://github.com/ethereum/wiki/wiki/White-Paper

[5] Dr Gideon Greenspan, *Multichain: Whitepaper* https://www.multichain.com/download/MultiChain-White-Paper.pdf

[6] December 15, 2016 *Storj: A Peer-to-Peer Cloud Storage Network* Shawn Wilkinson, Tome Boshevski, Josh Brandoff, James Prestwich, Gordon Hall, Patrick Gerbes, Philip Hutchins, Chris Pollard https://storj.io/storj.pdf

[7] Juan Benet *IPFS: Content Addressed, Versioned, P2P File System* https://ipfs.io/ipfs/QmR7GSQM93Cx5eAg6a6yRzNde1FQv7uL6X1o4k7zrJa3LX/ipfs.draft3.pdf

[8] JP Morgan, *Quorum: Whitepaper* https://github.com/jpmorganchase/quorum-docs/blob/master/Quorum-Whitepaper-20v0.1.pdf

[9] Muneeb Ali, Ryan Shea, Jude Nelson, Michael J. Freedman *Blockstack: A New Decentralized Internet* http://blockstack.org

Acknowledgment

This research work has been carried out under the leadership of the Institute for Technological Research SystemX, and therefore granted with public funds within the scope of the French Program Investissements d'avenir.