Document Validation using Blockchain
A validation scheme for natural person's documents

João V. Meyer[1] **, Lucas M. Palma**[1] **, Jean E. Martina**[1]

[1] Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) – Florianópolis, SC – Brazil

1994meyer@gmail.com, lucas.palma@posgrad.ufsc.br, jean.martina@ufsc.br

Abstract. *The great extension of Brazil's territory, combined with its demographics of more than 200 million inhabitants, results in a complex, slow and expensive notary system. Blockchain technologies can be of huge help in this scenario. It provides a decentralized peer-to-peer way of storing and validating documents. In this article, we start the discussion about a blockchain-based national notary system with means to store and validate the natural person's public documents. We prototype a solution, comprising of birth, marriage, divorce, and death documents/certificates. In the end, we present a comparison between the operational costs of the implemented prototype and the current notary system.*

Resumo. *A extensão do território brasileiro, combinado com sua demografia de mais de 200 milhões de habitantes, resulta em um sistema cartorário complexo, caro e lento. Tecnologias de blockchain podem ser de grande ajuda neste cenário. Elas nos provêm uma maneira distribuída de armazenar e validar dados em uma rede peer-to-peer descentralizada. Neste artigo, nós começamos a discussão sobre um sistema notarial nacional baseado em blockchain capaz de armazenar e validar registros públicos de pessoas naturais. Nós prototipamos uma solução que engloba os documentos de nascimento, casamento, divórcio e óbito. Ao final, são apresentados os custos operacionais do protótipo junto de uma comparação com o sistema cartorário utilizado atualmente.*

1. Introduction

The Brazilian notary system is composed of over 13 thousand institutions [Brasil 2019c]. They are privately controlled and have the legal power provided by the state [Rodrigues 2013]. Unfortunately, this vast network still uses archaic types of communication between its peers and paper to store most of its records. Furthermore, the Brazilian law states that every natural person's registered document is accessible by anyone interested in it [Brasil 1973]. We could say that the notary system is like a set of databases scattered throughout the Brazilian territory, each using its schema and formalities.

The problem arises when there is a need for communication between all of the institutions. Usually, this is a manual process. A person needs to make a request to another notary for a piece of document or information. This information flow makes the process very error-prone, slow, and expensive. A person makes mistakes, gets tired, misreads information, among other things. Besides, the possibility of losing documents or a whole a notary building in some catastrophe is real. Another problem is the scattering of data. Only notaries at the place of birth of some person are obligated to have the information about said birth [Brasil 1973].

There are already some works in the area. For instance, [Palma et al. 2019] proposes to issue higher education documents in a blockchain-based infrastructure. Other studies have discussed, in a similar manner, ways of using public blockchains as ledgers to authenticate university diplomas [Costa et al. 2018]. Also, in a more general discussion, a project proposes a protocol for notarizing documents inside a blockchain [Magrahi et al. 2018]. They do that in a way that guarantees achievability, retrievability, and proof of existence. Differently of the related works, our implementation focuses on the creation of a distributed data validation and storage network for notary records.

We organize this article as follows: In Section 2 we discuss some basic concepts of the Brazilian notary system and blockchain-based technologies; Section 3 presents, in details, the related works; Section 4 discusses the proposed solution for the problems described in this project; Sections 5 and 6 examine implementation and costs of operation in comparison with the current system.

2. Basic Concepts

Here we briefly present the building blocks of our proposal. In Section 2.1, we discuss public notaries and natural person's documents as they are the main object of our research. Section 2.2 presents the concepts of Blockchain, consensus protocols, and smart contracts.

2.1. Public notaries

In Brazil, public notaries are privately managed entities that have the power to authenticate and validate documents. Its power is delegated by the state [Rodrigues 2013, p. 232]. Law 6.015 from 1973, is the current law in practice which regulates notaries. It provided a model for registry books [Brasil 1973] defining how we should register such records.

Notaries have the attribution of providing authenticity to contracts, business, and all the parts involved in such actions. In this view, "notary law is the law of authenticity and format" (our translation) [Mustapich 1974]. When a notary authenticates a document, it officially recognizes it as a valid and truthful document [Pugliese 1989]. This interpretation is the main view used throughout this article.

2.2. Blockchain

A blockchain is a chain of blocks where each block contains a hash of the previous block. We can use this hash to verify the integrity of the previous block, i.e., to verify that the previous block was not changed. Moreover, we can make blocks of any data. The first published work that made use of Blockchain was in 1990 [Haber and Stornetta 1990]. It devised a way to create tamper-proof time stamps for computer files. The big difference between the modern blockchains and the one devised then is the decentralized network.

Every node in the network (a peer-to-peer network) has an identical copy of the Blockchain. We send every change performed in one of the nodes to every other node. Unfortunately, a big problem arises from it. How to guarantee that the changes made are valid/trustworthy? This question is famously known as the Byzantine Fault [Wensley et al. 1978]. To solve this problem, one can use consensus algorithms. For example, in the Bitcoin [Nakamoto 2008] network, nodes engage in Proof of Work (PoW) to record in the blockchain digital currency transactions. More precisely, a particular node call miner assembles Bitcoin transactions and store them in a new block in the Blockchain.

Moreover, the network applies monetary incentives, every time a miner creates a new block.

Another example of a blockchain platform is Ethereum [Wood et al. 2014], released in 2015. The main feature of Ethereum is its smart contracts. Nodes use a high-level programming language (e.g., Solidity) to write a so-called smart contract. We compile these contracts to byte codes that run in the Ethereum Virtual Machine (EVM).

Storing large quantities of data is expensive in Ethereum itself. However, developers usually avoid it by storing only a "link" to the data itself, for example, the hash of the data. This method saves considerable amounts of gas, which is Ethereum's way to charge for code execution in the network. The Interplanetary File System (IPFS) [Benet 2014] is a peer-to-peer distributed file system. IPFS creates a hash of every single file stored in it. The files are, subsequently, accessed using these same hashes. We often use it coupled with Ethereum as a way to store data related to a smart contract off-chain.

2.3. Smart contracts

Szabo, in 1997, coined the concept of smart contracts. He said, "we can embed contractual clauses (...) in the hardware and software we deal with", [Szabo 1997]. Today, it operates on the idea that if a blockchain network can agree on some random pieces of data added in the list of blocks, it can also validate computer programs' executions. As the execution is limited to only use data inside the chain itself, the execution of these computer programs should be the same in every node that has the same chain. We use this logic to validate smart contracts in the network. As a consequence, we permanently store every single state that the contract will ever have in the Blockchain.

We see smart contracts as computer programs that can execute arbitrarily complex logic. This property provides a new paradigm where participants do not need to rely on any third parties to "force" the execution of the contract's clauses (or a program). The whole network of nodes will act as the "enforcer". The Ethereum network is the most famous example of a smart contract-based blockchain. It provides a virtual machine that executes code, and charges for it, on every node of the network.

3. Related works

This section focuses on discussing related works that describe the uses of a blockchain as a way to store and validate documents in a decentralized manner. The first work is from [Palma et al. 2019]. The authors developed a validation scheme for higher education diplomas in Brazil, focusing on the degree certificate's issuance automation. A student who gets enrolled in a course has a predefined number of classes to be taken. As time goes by, the player adds finished classes to a contract representing the student's progress. When the student finishes the amount defined, the smart contract automatically issues a diploma. Unfortunately, our solution cannot make use of something predefined. Different from university courses, peoples' lives are not predefined. For instance, a person may, or may not, marry.

Moreover, we cannot predict when someone will marry, die, or be born, for that matter. Because of that, we cannot make an automated record emission system. Therefore, our proposal works as a set of records that is updated through time.

The second work, by [Costa et al. 2018], in a similar way to the previous one, discusses the use of blockchain technologies to register diplomas in a blockchain. Differently from our proposal and [Palma et al. 2019], this article does not perform any data validation on the registered documents. It uses the Blockchain as a simple ledger. It proposes creating publicly available APIs used by any entity that wants to authenticate a digital diploma. Our proposal could be adapted to work similarly, by creating a service layer between clients (a way for the user to interact with the data and contracts of our proposal) and the underlying blockchain service. Our proposal focuses on the blockchain logic, mainly smart contracts implementations.

The third and last, related work, defines a protocol for notarizing documents inside a blockchain [Magrahi et al. 2018]. Its functionalities are: documents archivability, retrievability, and proof of existence. It uses the Blockchain as a registry of actions in a trusted archiving solution. The data is stored mainly in the archiving solution, like a shared database, but only its metadata is stored in the Blockchain. This article is very similar to Provchain's [Liang et al. 2017] solution. It discusses a database solution where every interaction with it is registered in a blockchain network. Both works are essential to guarantee data provenience. However, unfortunately, they do not use the distributed data solutions available. IPFS, for example. Our implementation uses the distributed file system that IPFS provides and store non-essential data in it. Besides, our implementation focuses on data distribution and its validation.

4. Proposal

We propose to create a distributed record validation network using blockchain technologies. Each participant would publish records to it. As the data in a blockchain is replicated and distributed between its nodes, all participants would have access to all data stored in it, making it widely available and safely accessible.

First, we need to contextualize some definitions used here. A *record* is a set of data that represents a legal document. A *notary* is an abstract entity where records are inserted. In our implementation, we represent it by a smart contract. Figure 1 shows a basic, high-level view of the proposed system. Each notary, represented by the green icons, has some records (yellow) inserted into it. As the notary is a smart contract located in the blockchain network, all this data is stored in the Blockchain as well.

It is essential to highlight that our proposal has the following limitations: (a) Secrecy of records is not part of the scope of this proposal; (b) Every single record and data is considered public knowledge; (c) Security concerns regarding processes not blockchain-related, like user permissions, database accesses. In other words, the implementation shown here does not take into consideration Lei Geral de Proteção de Dados (LGPD) [Brasil 2018].

In the list above, for (a), there are legal situations in Brazil's law that some documents are not public knowledge and are kept secret from the public. Item (b) is related to item (a). All public documents, birth and marriage certificates, for example, are considered public and can be requested by anyone. Finally, item (c) states that we do not consider off-chain problems and limitations in our article. Possible hacks, steal/loss of credentials, among other things.

We need to set some requisites as well: (a) The designed system should be able to

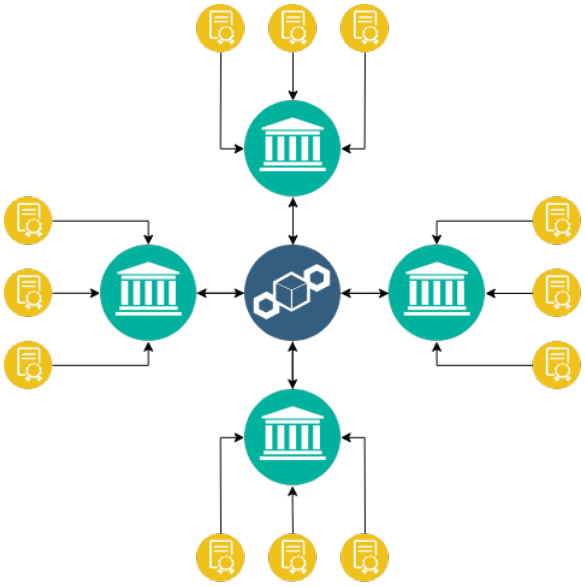

Figure 1. High level overview of the system.

incorporate existing documents; (b) The data in the Blockchain must have a way to link it to some other off-chain document

For (a), it is essential that we can migrate existing documents to this new system. In other words, we must incorporate past documents into it. Item (b) is concerned about linking data stored off-network with data inside it. We are concerned with this item for two main reasons: The first being that storing data in the Blockchain is expensive. It is a common practice to store information that can validate data on another, off-chain, location. A hash of the corresponding data, for example. Second, documents are usually created by different means: paper, virtual documents, and others. All of them containing the "source" data. Linking these documents with the records in the Blockchain makes the system more adequate.

5. Implementation

The prototype was implemented by partially mirroring functionalities of a physical notary where records are always related to a person. There are, also, the authorities of the notary. These participants are the only ones that have the legal power to register or edit records of a notary. We further explain each developed contract below. All the prototype code is available in the following URL (https://pastebin.com/zEcZ8sDH).

The *Notary* contract is one of the most straightforward contracts implemented. It consists of only two sets of information — one of the registered records and the other of the notary officials. The first one is used to store all the addresses of registered documents. If a document is registered, it is considered valid, just like its physical counterpart. On the other hand, we used the second set to store the addresses of people who have authority in the notary. We do this procedure to prevent unauthorized third parties from registering invalid documents.

The *Record* is where data is stored. This contract is an abstract contract and should be used with inheritance to create more specific ones. It contains a *validate* method that

Fact	Cost (BRL)
Birth	3.09154737
Death	0.9529527104
Marriage	1.256597943
Divorce	0.9815128428

Table 1. Cost of deployment of each smart contract developed

we use to validate data of a contract, and we implement it in every other subclass of *Record*. The data in the record makes it possible to validate it. For example, if the mother was not alive at the time of birth or if the father was not born at the time of birth, the document is invalid.

The *Person* contract represents an individual to which records refer. This contract is used as an anchor point to validate data from records. It contains references to possible contracts that a person may generate throughout his or her life. A birth; zero or more marriages; an equal number, or one less, of divorces compared to marriages; a death.

Furthermore, we conducted a partial cost evaluation of our prototype. The most important point is the deployment costs of our implementation. We obtained all costs listed here by deploying the contracts in a custom network using Remix IDE (a tool that helps developer code, test, and deploys smart contracts for Ethereum). All costs described here are considering the price of Ethereum on the day 19/10/2019, which was U$D 171.2 at 17:48, which is R$ 704.18. We represent all the transactions and executions costs by gas. The gas price was considered the base gas price of 1 gwei[1]. For every record registered, we also included the linking of an IPFS hash as a way to represent the linking between in chain data and real documents outside of it, according to the requisites section. Table 1 shows the result for all deployments.

According to Conselho Nacional de Justiça (CNJ), the declared revenue of Santa Catarina's notaries, in the first semester of 2019, was around 370 million Brazilian reais [Brasil 2019b]. Of these, about 95 million were spent in natural person registries. We have used the 95 million as a base to our calculations here because we built our prototype to mimic natural person documents only. We also double its value to consider a full year, totalling 190 million reais. It is valid to note that we could further develop the system implemented here to encompass all of the notary systems, but we are only concerned with this part for now. Of the 190 million reais, about 35% is considered operational costs [Luizari 2019]. This operational cost means that around 66 million reais are spent each year, only in the state of Santa Catarina, only with natural person documents, just to operate natural person's document registries.

If we consider the number of births, deaths, marriages, and divorces from Instituto Brasileiro de Geografia e Estatística (IBGE), we can draw exciting conclusions about the cost of running this prototype in large scale. Using data gathered from public government data sets [Brasil 2019a], we estimated the system cost in a whole year scale. Table 2 uses data from year 2017. It shows that, if 100% of the Santa Catarina state used our system, the cost would be meager compared to the current system.

[1]A gwei is one million Wei, which is the smallest fraction available for Ethereum transactions.

Fact	Occurrences	Total cost
Birth	98.978	R$ 305,995.18
Death	39.406	R$ 37,552.05
Marriage	34.098	R$ 42,847.48
Divorce	8.556	R$ 8,397.82
Total		**R$ 394,792.53**

Table 2. Yearly number of registries, by category, and costs.

Of course, our calculations do not take into account the costs of physical space, labor, among other things. We are only considering if the system was already in place, fully operational, its cost related to the blockchain network. Nevertheless, we could bring the value of 394 thousand reais further down. This value considers the price of Ethereum in the public network, which is expensive. Using a private network, like Quorum [Morgan 2016], could bring down costs to those of electricity and server maintenance.

6. Conclusion

As we can see, blockchain technologies are revolutionizing the way businesses, governments, and many other categories of industries work. It created a way of doing business logic without the need for a centralized point. Public institutions that operate using an accessible public blockchain can be audited by anyone, anywhere.

This article showed that estimated costs from our prototype represent just a fraction of the real cost of the Brazilian notary system. Also, when in place, it would reduce bureaucracy and make data easily accessible by everyone with an internet connection. However, implementing such a system has more details that do not fit the scope of this article. There are the problems of document secrecy, in cases of a judicial decision; Problems of adoptions, where we can not fully disclose the documents to anyone; And others. These are only the practical problems. The biggest challenge would be legal since Bitcoin has popularized the term "blockchain". Unfortunately, lawmakers and the population, in general, do not understand it fully. The media usually draws attention to its bad uses, such as ransomware. These problems make the acceptance process a prolonged and difficult one.

However, economic issues usually prevail in these cases. As this study showed, the use of Blockchain for notary systems could improve the quality of service and bring new, unseen, features to it using a highly distributed, public, and freely accessible data set of public records. This implementation is only for notaries, but we could use it in many more sectors. Government contracts, transparency, elections are just some of the possibilities.

References

Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561*.

Brasil (1973). Lei nº 6.015 de 1973. "`http://www.planalto.gov.br/ccivil_03/Leis/L6015compilada.htm`". [Online, accessed on 26/02/2019].

Brasil (2018). Lei nº 13.709 de 2018. "http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm". [Online, accessed on 19/10/2019].

Brasil (2019a). Ibge. "https://www.ibge.gov.br/estatisticas/sociais/populacao". [Online, accessed on 19/10/2019].

Brasil (2019b). Justiça aberta. "https://www.cnj.jus.br/corregedoria/justica_aberta". [Online, accessed on 19/10/2019].

Brasil (2019c). Lista de cartórios do brasil. "http://dados.mj.gov.br/dataset/lista-de-cartorios-do-brasil". [Online, accessed on 19/06/2019].

Costa, R., Faustino, D., Lemos, G., Queiroga, A., Djohnnatha, C., Alves, F., Lira, J., and Pires, M. (2018). Uso não financeiro de blockchain: Um estudo de caso sobre o registro, autenticação e preservação de documentos digitais acadêmicos. 1(1/2018).

Haber, S. and Stornetta, W. S. (1990). How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography*, pages 437–455. Springer.

Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., and Njilla, L. (2017). Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In *Proceedings of the 17th IEEE/ACM international symposium on cluster, cloud and grid computing*, pages 468–477. IEEE Press.

Luizari, L. (2019). Repasses e despesas: Para onde vai o dinheiro pago aos cartórios brasileiros? *Cartórios com Você*, (8):25.

Magrahi, H., Omrane, N., Senot, O., and Jaziri, R. (2018). Nfb: A protocol for notarizing files over the blockchain. pages 1–4.

Morgan, J. (2016). Quorum whitepaper. *New York: JP Morgan Chase*.

Mustapich, J. M. (1974). *Revista Notarial Brasileira - nº. 1*.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Palma, L. M., Vigil, M. A., Pereira, F. L., and Martina, J. E. (2019). Blockchain and smart contracts for higher education registry in brazil. *International Journal of Network Management*, page e2061.

Pugliese, R. J. (1989). *Direito Notarial Brasileiro*.

Rodrigues, M. G. (2013). *Tratado de registros públicos e direito notarial*. Editora Atlas SA.

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday*, 2(9).

Wensley, J. H., Lamport, L., Goldberg, J., Green, M. W., Levitt, K. N., Melliar-Smith, P. M., Shostak, R. E., and Weinstock, C. B. (1978). Sift: Design and analysis of a fault-tolerant computer for aircraft control. *Proceedings of the IEEE*, 66(10):1240–1255.

Wood, G. et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. 151:1–32.