Federal University of the Rio de Janeiro State

Center of Exact Sciences and Technology

School of Applied Informatics

Universidade Federal do Estado do Rio de Janeiro

Centro de Ciências Exatas e Tecnologia

Escola de Informática Aplicada

Decentralized Student National Identification: A Blockchain Approach

Lucas de Souza Ribeiro

Supervisor

Pedro Nuno de Souza Moura

Rio de Janeiro, RJ – Brazil

January 2019

Catalogação informatizada pelo autor

R484 Ribeiro, Lucas de Souza
Decentralized Student National Identification: A Blockchain Approach / Lucas de Souza Ribeiro. -- Rio de Janeiro, 2019.
27

Orientador: Pedro Nuno de Souza Moura.
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Estado do Rio de Janeiro, Graduação em Sistemas de Informação, 2019.

1. Blockchain. 2. Meia-Entrada. 3. Identificação. 4. Estudante. I. Moura, Pedro Nuno de Souza, orient. II. Título.

Decentralized Student National Identification

Lucas de Souza Ribeiro

Undergraduate thesis presented to the Applied Informatics School of the Federal University of the Rio de Janeiro State to obtain the Bachelor Degree in Information Systems.

Approved by:

_____

Pedro Nuno de Souza Moura (UNIRIO)

_____

Leonardo Luiz Alencastro Rocha (UNIRIO)

RIO DE JANEIRO, RJ – BRAZIL.

JANUARY 2019

Acknowledgements

Since I was a kid I had a dream to complete my graduation. After many years it is finally getting realized and all I can do now is thank every teacher that helped me to make it possible, but mostly the UNIRIO for being such a great educational institution, and also my brother and my mom for supporting me.

RESUMO

Em conjunto a lei da meia-entrada baseada nas carteiras estudantis surgiram também atividades fraudulentas que afetam diretamente o funcionamento da mesma. O problema se agrava com a ineficiência da autenticação, que viabiliza a criação de cópias dos códigos autenticadores e sua distribuição, além da aceitação de um única entidade onde a transparência dos dados também é precária para posterior auditoria. Considerando que cada instituição tem suas próprias regras de geração da identificação estudantil e verificação da mesma, a blockchain é uma tecnologia recente que apresenta uma solução através de um sistema distribuído. Aplicada a este caso possibilita a autonomia das instituições e uma verificação única processada pelo apoio da carteira de identificação nacional.

Palavras-chave: blockchain, meia-entrada, identificação, estudante.

ABSTRACT

Along with the “Lei da meia-entrada”, or middle-entry law as its literal translation, based on student identifications also appeared fraudulent activities that directly affect its operation. The problem is aggravated by the inefficiency of authentication, which makes it possible to create copies of the authenticator codes and their distribution, in addition to the acceptance of a single entity where the transparency of the data is also precarious for subsequent auditing. Considering that each institution has its own rules for generating students identification and verification, blockchain is a recent technology that presents a solution through a distributed system. Applied to this case allows the autonomy of the institutions and a unique verification processed by the support of the national identification card.

Keywords: blockchain, middle-entry, identification, student.

Index

Índice de Figuras

1 Introduction

1.1 Motivation

After some years of fraudulent activities concerning fake student identities in Brazil for having almost no authenticity control, resulting in consequences for the entertainment and culture industry, then it has been released a new law to create a national pattern in late 2013 trying to solve these problems. But along with the new law we got issues like a huge bureaucracy and even a cost to be paid by students to make use of their own rights.

As expected there is a limited time for the validity of the new ID, starting from 1 April to 31 March of the next year, but under the student perspective he should pay to use his right for a limited amount of time, which already enables improvements. Consider also that the value per time of validity gets more expensive as time passes, along with the waste of resources to produce a physical card and to ship the product to the student (also increasing costs).

According to the 2016 census presented in (INEP, 2017) disclosed by the *Ministério da Educação* (MEC) and *Instituto Nacional de Estudos e Pesquisa Educacionais Anísio Teixeira National* (INEP) there were 8.052.254 students enrolled in 2.407 universities, if they expected to reach every single one using this law enforced student identification, considering the value of 35 BRL for each card, thats a cost of over 280 million BRL, some of the students under specific circumstances can be released from this cost, but this amount is not even considering the shipping costs.

We can suppose that under the cost, bureaucracy and time enrolled to a student get this card every year, this provided solution do not reach a substantial percent of the total students. This new law can also be considered flawed in some aspects, when trying to be more accessible by not specifying strictly which institutions could provide the identification card, still allowing some fraudulent activities as the law-enforced control

for these institutions get much harder. The possibility to rely on a digital, and more cryptography dependant solution could be cheaper, faster and more secure.

1.2 Objectives

With the continuous growth of blockchain technology, it has the characteristics to achieve a decentralized system that every university could have autonomy to provide their own data of active students. While they could authenticate every information which is inserted inside the blockchain, and publicly provide a get request to allow everyone check if an information is present into the blockchain.

The technology itself provides lots of security and control aspects that could be used for the success of this project, while making some tweaks to adapt the consensus of the restricted submissions and the validity check.

This way we can provide such a system to allow a student go to wherever he normally does when looking for entertainment, while carrying his National ID card as expected, and just present it to who is responsible to check if he could get a discount for being a student under the law. In a matter of seconds using a cellphone or any type of device connected to the internet, that supports an application able to request a validity operation into the blockchain, his identification present into the National ID called *Cadastro de Pessoa Física* (CPF) can be verified as active into the brand new national student register.

1.3 Text Structure

The present work is structured in the following chapter organization beyond the introduction:

Chapter 2: Defines the main concepts that need to be understood before the presentation of the solution, such as the law, the blockchain technology and the elliptic curve digital signature algorithm.

Chapter 3: The solution is presented as a conceptual solution proposal of the distributed system using the Blockchain.

Chapter 4: Sample implementation of how the validation should work.

Chapter 5: Present the final considerations, limitations of study, identified issues and suggestions towards the next step to achieve a fully working system.

2 Main Concepts

2.1 Law 12.933/13 & Decree 8.537/15

The Law 12.933/13 defined in (BRASIL, 2013), define the benefit of access to artistic, cultural and sporting entertainment for half the price under a few circumstances for students, old people, disabled people and young people aged 15 to 29. This law defined that the student could only prove his condition by presenting the Student Identity Card named as *Carteira de Identificação Estudantil* (CIE) and applying emission restrictions to only specific regulamented organizations.

Enforcing this law, the Decree 8.537/15 defined in (BRASIL, 2015) determined the new adopted legal pattern of the CIE making use of common and national accepted information present in documents such the National ID with additional information about the university the student is enrolled at and the validity until 31 march of the subsequent year of expedition.

Besides the law enforces this restriction to allow only some regulamented institutions to issue this identification, while some of these organizations could be easily accessible to students, the pattern itself makes some barriers in need of all the infrastructure to manufacture while following the padronization, even considering that there are specific organizations like Academic Directories and Central Student Directories which are linked to university students, that do not receive the proper investments even when inside public institutions, to make it possible.

2.2 Blockchain Technology

The Blockchain approach made its public debut in (NAKAMOTO, 2008), where it was provided a solution for a decentralized payment system which disposes a public ledger replicated over nodes in the network, using algorithms to secure that every node

works in an up-to-date version of the ledger, while maintaining every transaction divided in blocks, which are linked together by the hashes of the previous block, then using an algorithm to provide a consensus between all over the network to decide which blocks are valid and must be attached to the chain.for the Blockchain, as appointed by Satoshi Nakamoto after k number of blocks, the

The “Proof of Work”, an algorithm proposed in (DWORK and NAOR, 1992) to combat junk mail, made a vital importance on the first application of Blockchain, to make the technology secure against the “Double Spending” issue and fully public at the same time, every block must be hashed and its resulting value must be under a target to be considered a valid input. The effort needed to change any data inside of a single block, hash it and also hash all the subsequent blocks while disputing with all the network on the other hand, becomes exponential.

This proof is not alone the only possibility to achieve a consensus, another common one is the “Proof of Stake” which is getting a great visibility as it do not require that much energy and processing, delegating the power of decision proportionally to the higher influencers while they are also the ones who have the most to lose if the system is compromised.

Some blockchain implementations could also be restricted in submissions, they are generally called private or hybrid, not everyone can post a block into the chain, the validity check crosses over rules that only specific requests can be approved, as for example, using a digital signature under the blockchain knowledge can secure that one is capable of being accepted under the validation and others cannot.

2.3 Elliptic Curve Digital Signature Algorithm

When information is transmitted between parties, the relevance of this information induces the recipient to know if the integrity of the message has not been compromised, and also if the sender is really the owner.

The Digital Signature Algorithm (DSA) was proposed in 1991 and later adopted as a standard stated in (FIPS, 1994), following various digital signature schemes. As

signatures it works like a proof that the document was acknowledged by the one who signed it, so does its digital analog, but the issue is that digital documents could be easily faked to state this information. So here the algorithm comes by, it can create a key defined by a string of characters, and process a message along with this key to generate a signature in bytes.

This key allow the generation of another key from it, which cannot lead to the first one, the first is commonly called Signing Key or Private Key, and the second one called Verifying Key or Public Key. They are a key pair, it means that the Signing Key can sign a message and only the Verifying Key generated from that key could verify if the message and the signature are authentically provided by its Signing Key.

As an example, if a message and a Verifying Key shows true for a signature, its proven that only the owner of the Signing Key that generates this specific Verifying Key can be the owner of the signed message. It also shows why they can be defined as a private and public, as the Private Key leads to the owner of the signature, and the Public Key can be distributed to anyone who the owner wants to just check the authenticity of the message.

One of the DSA famous variants is the Elliptic Curve Digital Signature Algorithm (ECDSA) defined in (FIPS, 1998), which uses the structure of elliptic curves to achieve a better security level under smaller sizes, it is also used in the Bitcoin Protocol.

3 Solution

The solution proposed by this work is a kind of private or also called hybrid blockchain, where the more common proofs of consensus are not advantageous as we do not need a huge work to ensure validation but instead distribute the power of decision in equality to every node, as the autonomy of every institution.

The idea is divided in two blockchains seen in Figure 1 below, one called DSNI Blockchain, which stores the students identifications that should allow public access for validation requests, and the PKEY Blockchain, responsible to store the public keys associated with every institution. Each of these Blockchains have their network accepting wide communication inside their own, but both should only communicate with each other in a local way, as the PKEY Blockchain must have access restrictions since its data is not relevant outside the main blockchain verifications.

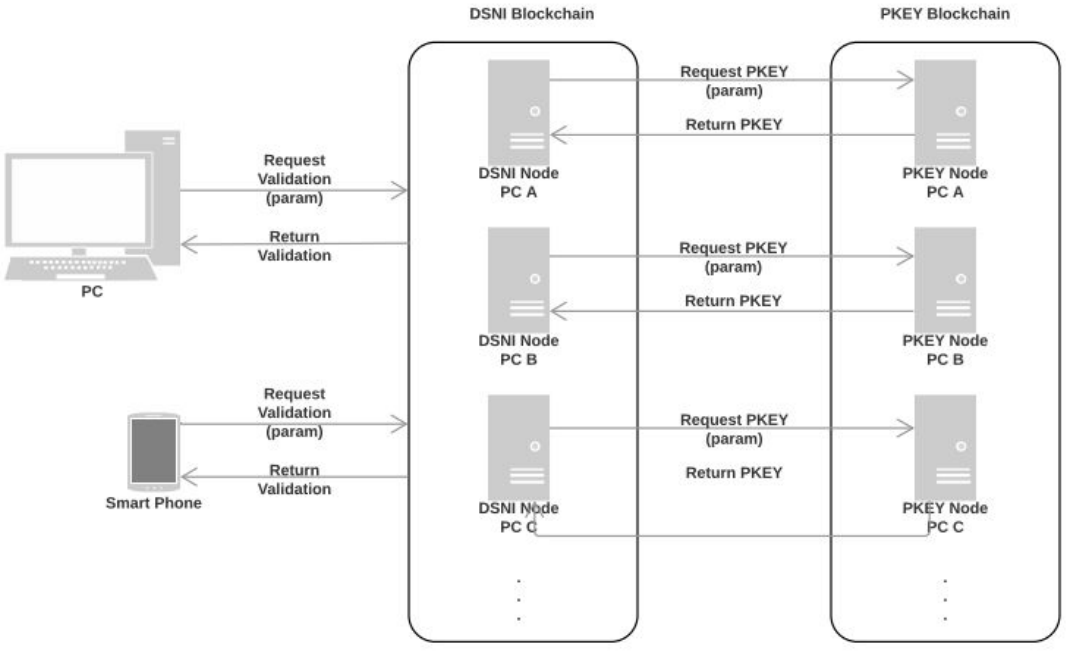

Figure 1. Blockchains communication representation.

A single computer in each university run a node of each blockchain, with a front application to allow an user to input the needed informations to form a block submission to the DSNI Blockchain in case of registering student identifications, or PKEY Blockchain if its about providing authority to another institution.

Each block submitted to the chain is verified under the Proof of Authority if its owner have the provided submission permission inside the PKEY.

For example, an institution can send its students identifications to the chain, registering every CPF inside a block, signing its sensitive data with its private key and submitting to the DSNI. Then the blockchain will check if every information required is provided, request from the PKEY, the public key that can verify the block, sending any type of identification, and then checking if the message, signature and public key confirms for a valid entry. If yes the block get appended to the chain, otherwise gets rejected.

Beyond this verification there are some organizational issues that must be identified, as only nodes of the blockchain could participate actively in validations and submissions. Devices from outside could only use it to request if a CPF is valid, even a batch request for entertainment organizations could also be possible.

The desired public to attain with this solution is counting only universities, so as stated before, around 8 million CPFs and 2.400 nodes. In a single year every node should have around 2 submissions, one per semester, from each node, possibly more if there are any errors or delayed students. This could roughly generate 100 megabytes per year of data, considering the full block structure and that CPFs are 11 digit numbers being the most of this consumption. It can shorten the duration of the student identification to be between 6 to 8 months, allowing a better control.

Chain validation should be needed at every verification, it can be costly and still has to be evaluated. On the other hand the consensus of the adopted chain into the network could be applied daily as this synchronization does not need to be much faster. But still need a proper way to apply this consensus, as the current longest chain treatment cannot be used, simply because a single node with authority could submit the number of blocks plus one to its own blockchain, and erase the rest of the network spreading its own chain.

The blockchain structure provides the needed that all the present information could not be erased easily, creating an easy way to audit if there are any incorrect information being submitted as the responsible is identified.

If any possible attacker have access to an authorized private key, he could only submit blocks under the proper institution identification, so it remains traceable. The content of the submitted blocks could be used improperly for some time, but the amount of effort is questionable. For this reason the PKEY Blockchain should only iterate the values backwards, looking for the most recent public key allowing to forget compromised keys.

4 Implementation

Using python as the programming language, along with support libraries allowing a simple use of algorithms, json communication, testing and execution environments to achieve a sample of how this blockchain could be implemented. Also expanding the idea through this work, pointing possibilities and identifying flaws for further improvement.

4.1 DSNI Blockchain

As the Blockchain is literally a data structure based in Blocks, as the first step, the block must be defined as it is showed below in Figure 2.

```
block = {'index': len(self.chain) + 1,
         'timestamp': str(datetime.datetime.now()),
         'signature': signature,
         'data': {
                 'previous_hash': previous_hash,
                 'identifications': identifications,
                 'institution_id': institution_id
                 }
}
```

Figure 2. Code Representation of DSNI Block Structure

- **Index**: Define the length of the blockchain, while also creating a numerical link between the blocks.
- **Timestamp**: Vital to store a chronological link, also being used to create a barrier to validate the order of the accepted blocks when there is any kind of conflict and serves a purpose to identify the validity of the identifications present in each block submission, as it can be considered expired after some time.
- **Signature**: Hashed data information signed by the ECDSA algorithm under a Private Key that can assure a verification of the authority along side with the

data when using the designed Public Key generated from that specific Private Key who signed the message data.

- **Data**: Where the information is stored and afterwards hashed and signed for the sake of verification.
    - **Institution_id**: ID responsible to link the authority of the block, since it defines which institution made this submission and is the key to achieve the public key designed to solve the verification of the block.
    - **Identifications**: An array containing all the CPFs provided by the institution of every active student at the moment of the block submission.
    - **Previous_hash**: Main reason how the blockchain is considered chained, as every block must carry the hash of the previous block, guaranteeing that each block has a correct position on the chain, right before and after a certain block.

These fields carry a lot of information, but still need some treatment from algorithms to effectively work as a blockchain. On the next step the data structure must be refined to respect the integrity of the chain and to attend some simple rules so as to create a basic sense of security.

Timestamp and index fields are generated by the receiver node, as the index must respect the actual length of the chain without any other purpose, and the timestamp must only be used to verify how old a block and its data are.

The remaining values are the data provided by the post request and its signature. Inside the “data” field there is the “institution_id” which creates a link with the side PKEY Blockchain responsible to provide the Public Key that can verify the authenticity of this data using also the signature field, identifying the called Proof of Authority. The “identifications” field to make sure that the core information cannot be modified and the “previous_hash” that guarantees the position of the block. So even the same block could not be resubmitted even from any other party, as it would be rejected since the hash of the previous block certainly will not be the same.

Under the protection of the hashing algorithm, this proof can secure that not even a single character could be modified, inside any of these fields, since only the owner of the accepted key pair could recreate the authenticity. This information being saved inside the blockchain, allows for a history of which institution sent each block. Where the institution code could not be different from signing owner of the message, as this information has to be checked on the PKEY Blockchain.

4.2 PKEY Blockchain

When looking at an authority based proof, there are some aspects that must be attended:

- ❏ Authorized requesters must have its Public Key available to each node into the network.
- ❏ Only Public Keys from authorized requesters should be available to the nodes to verify.
- ❏ The set of Public Keys must be only incremental, as any added block could be verified as valid for its lifetime and if a Private Key is compromised, a new Public Key from its authority must be added.

Instead of building a repository, a blockchain can provide a secure implementation of every aspect above, using Proof of Authority. This creates a loop, creating another blockchain to provide the information to check the authority. To solve this problem it reaches a slightly downside, that creates an initial centralization to distribute the authority. Basically only an already authorized existing node could submit a Public Key to allow the paired Private Key owner as a new accepted authority.

```
block = {'index': len(self.chain) + 1,
         'timestamp': str(datetime.datetime.now()),
         'signature': signature,
         'data': {
                 'previous_hash': previous_hash,
                 'public_key': public_key,
                 'institution_id': institution_id,
                 'registrant_institution' : registrant_institution
                 }
}
```

Figure 3. Code Representation of PKEY Blockchain

Figure 3 shows that this block uses the same idea of the structure in DSNI, the main data is now the “public_key” field, and the new “registrant_institution” allows to identify which institution allow this Public Key to be a part of the authority system, and it is directly linked to the verification of its own block. Only the institution that is registering this block can be the owner of this block, as it should iterate over the possible keys into the chain, until it reaches the desired “institution_id” and provide its “public_key”.

4.3 Functions

With the block structures already defined, there are some important web methods to point out.

```
function submit_pkey();
        SET json to request.get_json()
        SET submission_keys to (‘data', 'previous_hash', 'public_key', 'institution_id',
'signature', 'registrant_institution')
        FOR EACH key in json
                IF submission_keys.contains(key)
                REMOVE key from submission_keys
        IF submission_keys is not empty
                RETURN “Some elements of the transaction are missing”
        ELSE
                SET block to create_block (signature, previous_hash, public_key,
institution_id, registrant_institution)
                SET proof to check_block_poa(block)
                IF proof is equal True
                        RETURN “Block Authority Confirmed! Block added to the chain”
                ELSE
                        RETURN “Authority not confirmed! Block rejected ”
```

Figure 4. Submit Public Key post request method pseudocode

After genesis blocks of both blockchains have been created, the PKEY Blockchain must be initialized with an authorized Public Key, so the initial owner could spread this authority to other institutions using the example method in Figure 4. Starting with a verification if all obligatory fields are received. If yes, the new block is created using the provided information and sent to Proof of Authority verification seen in Figure 5 below.

```
function check_block_poa(block) returns boolean;
    SET proof to False
    SET previous_block to lastItem from chain
    SET previous_hash to hash(previous_block)
    IF previous_hash is not equals block.previous_hash
        RETURN False
    SET registrant_institution_id to block.registrant_institution
    SET authority_pkey to get_registrant_pkey(registrant_institution_id)
    SET proof to authority_pkey.verify(signature, block.data)
    IF proof is equal True
        ADD block to chain
        RETURN True
    ELSE
        RETURN False
```

Figure 5. Check block Proof of Authority method pseudocode

Since new Public Keys are submitted and accepted by the proof, using the registrant institution of the genesis block and its Public Key to verify. Now the DSNI Blockchain can start expand, getting block submissions from every authorized institution under the PKEY Blockchain, achieving almost the same methods as shown in Figures 4 and 5.

In DSNI Blockchain the submission is called “submit_block”, doing the same workflow, until the “check_block_poa” of this Blockchain instead of iterating over its own chain, creates a post request expecting the Public Key of the sent “institution_id”. With the response being able to verify the authority of the request and possibly append the new block its chain.

Some other methods were created to imply in a distributed format of the blockchain, allowing communications between nodes, to replace the chain, but the consensus of the chain replacement is using the longest chain. This is not suitable for the use of Proof of Authority, as a single node could take over the network chain by submitting multiple blocks until it surpasses the length of the network blockchain.

5 Conclusion

5.1 Final Considerations

Through the study of this work, the idea to renew the way of identifications in a decentralized system of organizations such as universities made a lot of progress aiming to adapt to such difficulties encountered, even though some were still created and have to be dealt with.

The scalability of this is proposed at first to all universities of Brazil considering the current law of Rio de Janeiro state, where young people below 21 years old have the right of the middle-entry law just by presenting the national identification card. And at this level of education presumes that the institutions have the capacity to deploy a single node and can take the appropriate measures to secure the sensitive information needed.

It is important to emphasize that this accessibility aided by the law is common in Brazil, but it has created an unequal treatment as it is expected that the children and youth depends on other people assistance to pay for cultural events. But the industry on the other hand raised prices as the fraudulent activities began to heavily impact on their profits.

5.2 Study Limitations

The size of the object achieved by this study is enormous when looking at all the disciplines involved and the deep knowledge needed to approach every aspect that could be considered an issue to possibly develop a solution. Network, connection of the nodes and lot of security issues were not heavily discussed based on the time needed and personal limitations.

5.3 Future Works

A solid base of information have been created towards the objective, but still some very important issues have come by, the consensus to assume a chain between a conflict cannot use a longest chain rule. It does not work for Proof Of Authority system without at least some changes that still has to be identified, or even another possibility of consensus that could fit for this particular way of implementation still have to be found.

The centralization of authority distribution is a problem for the initial stage of the blockchain, but as a blockchain system can be auditable and a solid implementation should make it immutable for the long run.

Optimization of this system has not been evaluated yet, the idea is based on an hypothetical deployment to use every university of Brazil as a node creating around 2.400 nodes to communicate to each other, and responding to get requests. Having to iterate through the chain as it is the way to guarantee that the information is secured by the blockchain structure, also should be considered balanced distribution of external requests, looking for proximity, power of computation and traffic control.

References

ANSI. **Public Key Cryptography For The Financial Services Industry: The Elliptic Curve Digital Signature Algorithm (ECDSA)©**. 1998. Available in: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.2977&rep=rep1&type=pdf>. Accessed in: 20 nov. 2018.

BRASIL. **DECRETO Nº 8.537, DE 5 DE OUTUBRO DE 2015**. Available in: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Decreto/D8537.htm>. Accessed in: 28 sept. 2018.

BRASIL. **LEI Nº 12.933, DE 26 DE DEZEMBRO DE 2013**. Available in: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2013/Lei/L12933.htm>. Accessed in: 28 sept. 2018.

DWORK, C.; NAOR, M. **Pricing via Processing or Combatting Junk Mail**. 1992. Available in: <https://web.cs.dal.ca/~abrodsky/7301/readings/DwNa93.pdf>. Accessed in: 29 oct. 2018.

FIPS. **Digital Signature Standard (DSS)**. 1994. Available in: <https://web.archive.org/web/20131213131144/http://www.itl.nist.gov/fipspubs/fip186.htm>. Accessed in: 20 nov. 2018.

INEP. **MEC e Inep divulgam dados do Censo da Educação Superior 2016**. 2017. Available in: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/mec-e-inep-divulgam-dados-do-censo-da-educacao-superior-2016/21206>. Accessed in: 27 nov. 2018.

NAKAMOTO, S. **Bitcoin: A Peer-to-Peer Electronic Cash System**. 2009. Available in: <https://bitcoin.org/bitcoin.pdf>. Accessed in: 29 oct. 2018.