MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA

Universitatea Tehnică a Moldovei

Facultatea Calculatoare, Informatică și Microelectronică

Departamentul Inginerie Software și Automatică

Admis la susținere
Şef departament: Fiodorov I. dr., conf. univ.
____
„____" _________________ 2022

Aplicarea tehnologiilor blockchain și IPFS în asigurarea confidențialității și accesibilității datelor sensibile

Teză de master

Student:	_______________	**Dodon Ion, gr. IS-211M**
Coordonator:	_______________	**Zaharia Gabriel, asist. univ.**
Consultant:	_______________	**Catruc Mariana, lect. univ.**

Chişinău, 2023

REZUMAT

Keywords: Date, securitate, decentralizare, blockchain, IPFS.

Domeniul de tehnologii informaționale se dezvoltă foarte repede si apar tot mai multe tehnici și tehnologii noi ce fac acest domeniu să progreseze. Unele dintre cele mai noi idei apărute in ultimii ani sunt de a folosi principiile democratiei pentru a rezolva anumite probleme in era noastră informațională. Aceste probleme fiind legate de siguranța datelor, despre recunoașterea apartenentei lor, si altele.

Această lucrare sub numele "Aplicarea tehnologiilor blockchain și IPFS în asigurarea confidențialității și accesibilității datelor sensibile" elaborată în cadrul tezei de master, de către studentul Ion Dodon al Universității Tehnice a Moldovei, are ca scop de a rezolva unele probleme ce țin de confidențialitatea datelor sensibile. Si pentru aceasta se vor folosi tehnologiile blockchain si IPFS. Acestea având caracteristicile necesare pentru a crea un sistem sigur de protecție a datelor care nu depind de o companie terță.

Lucrarea este împărțită in următoarele părți: analiza domeniului, specificarea cerințelor funcționale și mai ales non-funcționale, design-ul sistemului, si implementarea lui. Iar la sfârșit este concluzia unde se menționează despre observațiile făcute pe parcursul lucrării și rezultatele obținute.

Internetul are o importanță din ce in ce mai mare pe zi ce trece. In ficare zi, in fiecare arie din societate sunt folosite calculatoarele personale, servere, telefoane mobile inteligente si toate acestea comunică între ele. O mare parte din datele despre noi, care ne expun personalitatea, sunt stocate in mediul on-line. Aceste date pot fi folosite cu scopuri rele, de exemplu ele pot fi colectate de o companie sau chiar persoane individuale fiind vândute altor companii sau persoane individuale pentru ca mai târziu societatea sa fie manipulata deja fiind cunoscută destul de bine ce preferințe are sau alte alte date despre populație. Multe din datele stocate pe servere sunt de natură sensibilă, cum ar fi acte personale medicinale. Desigur la moment aceste date sunt păstrate cât de sigur posibil, dar pană la urmă ele sunt pastrate pe servere care aparțin unor companii ce ofera servicii IT (în cloud). Ceea ce înseamnă că noi incredințăm lor aceste date. Alt aspect de risc ar fi cazul când aceste servere cad si datele nu mai pot si recuperate. Problemele mentionate sunt bine cunoscute si există măsuri de prevenire, dar sigur că aceste măsuri nu sunt perfecte. Această lucrare iși propune sa vină cu noi idei de a proteja datele utilizatorilor de internet fără a depinde de o companie externă. Aceste date care să nu le poată obtine nimeni altcineva decât deținătorul lor, si care nu se pot pierde în cazul cand un server cade.

Pentru a insuși acest nivel de securitate va fi creat un sistem care stocheaza datele pe o rețea IPFS ceea ce înseamnă că va fi un sistem decentralizat si în cazul cand un server cade, datele sunt repede recuperate pe alte noduri (servere). Va fi creat un Smart Contract care va functiona ca un manager de date și va memora cine și ce date deține. Toate aceste operații vor fi transparente și datele stocate for fi encriptate.

ABSTRACT

Keywords: Data, security, decentralization, blockchain, IPFS.

The field of information technologies develops very quickly and more and more new techniques and technologies appear that make this field progress. Some of the newest ideas that have emerged in recent years are to use the principles of democracy to solve certain problems in our information age. These problems are related to the safety of the data, about the recognition of their belonging, and others.

This work under the name ”The application of blockchain and IPFS technologies for assuring confidentiality and accessibility of sensitive data” elaborated as part of the master’s thesis, by the student Ion Dodon of the Technical University of Moldova, aims to solve some problems related to confidentiality sensitive data. And for this, blockchain and IPFS technologies will be used. These having the necessary features to create a secure data protection system that does not depend on a third-party company.

The work is divided into the following parts: domain analysis, specification of functional and especially non-functional requirements, system design, and its implementation. And at the end is the conclusion where it is mentioned about the observations made during the work and the results obtained.

The Internet is becoming more and more important every day. Today, personal computers, servers, and smart mobile phones are used in every area of society and all of these communicate with each other. A large part of the data about us, which expose our personality, is stored in the online environment. This data can be used for bad purposes, for example, it can be collected by a company or even individuals being sold to other companies or individuals so that later the company can be manipulated already knowing quite well what its preferences are or other data about the population. Much of the data stored on the servers is of a sensitive nature, such as personal medical records. Of course, at the moment these data are kept as safe as possible, but until the end, they are kept on servers belonging to companies that offer IT services (in the cloud). Which means we entrust them with this data. Another aspect of risk would be the case when these servers fall and the data cannot be recovered. The mentioned problems are well known and there are preventive measures, but surely these measures are not perfect. This paper aims to come up with new ideas to protect internet users’ data without depending on an external company. These data cannot be obtained by anyone other than their owner, and that cannot be lost if a server goes down.

In order to acquire this level of security, a system will be created that stores data on an IPFS network, which means that it will be a decentralized system and in case a server falls, the data is quickly recovered on other nodes (servers). A Smart Contract will be created that will act as a data manager and remember who owns what data. All these operations will be transparent and the stored data will be encrypted.

Contents

List of figures . . . 6

List of tables . . . 8

Acronyms . . . 9

Listings . . . 10

INTRODUCTION . . . 10

1 DOMAIN ANALYSIS . . . 11

1.1 Problem analysis and definition . . . 11

1.2 Scope and solution . . . 12

1.2.1 The inspiration . . . 12

1.3 Importance of the topic . . . 13

1.4 Existing solutions . . . 14

1.4.1 IPFS pinning services versus Google Drive . . . 14

1.4.2 Advantages and disadvantages . . . 16

1.4.3 Cost analysis . . . 19

1.5 Proof of Concept . . . 20

2 REQUIREMENTS SPECIFICATION . . . 24

2.1 Functional Requirements . . . 25

2.2 Non-Functional Requirements . . . 26

3 SYSTEM DESIGN . . . 28

3.1 Use cases . . . 28

3.2 The environments involved . . . 30

3.3 High level architecture . . . 30

3.4 Data protection using AES . . . 31

3.5 Data storage flow . . . 33

3.6 Integrated components . . . 35

3.6.1 MetaMask . . . 35

3.6.2 Etherscan . . . 36

3.6.3 IPFS Desktop . . . 37

3.6.4 Blockchain testnets . . . 37

4 SYSTEM IMPLEMENTATION . . . 39

4.1 Technologies . . . 39

4.1.1 Solidity 39

4.1.2 Hardhat 40

4.1.3 IPFS-core 41

4.1.4 Data protection using AES 41

4.1.5 Goerli testnet 42

4.2 The Smart Contract 44

4.2.1 Build process 44

4.2.2 Resulting artifacts 51

4.2.3 Deployment 53

4.3 The client application 54

4.4 License 56

CONCLUSIONS 58

Bibliography 60

Appendices 61

A The Smart Contract 61

List of Figures

1.1 Comparing the movement of data in IPFS to centralized client-server models 13
1.2 Google Drive Pricing 19
1.3 Smart Contract operations consuming Gas 21
1.4 Gas refund process 22
1.5 Solidity operations gas usage 22

2.1 Types of requirements 24

3.1 Main use cases 29
3.2 High-level architecture 31
3.3 Symmetric key cryptography 32
3.4 Asymmetric key cryptography 33
3.5 Data storage flow 34
3.6 Smart Contract development and the related components 35
3.7 MetaMask extension interface 36
3.8 Transaction history listed in Etherscan 37

4.1 Summarised comparison between DES and AES 42

Acronyms

ABI Application Binary Interface.

AES Advanced Encryption Standard.

API Application Programming Interface.

CID Content Identifier.

DES Data Encryption Standard.

EVM Ethereum Virtual Machine.

GUI Graphical User Interface.

IPFS Inter Planetary File System.

JSON JavaScript Object Notation.

NFT Non-Fungible Token.

NPM Node Package Manager.

P2P Peer-to-peer protocol.

PoS Proof of Stake.

PoW Proof of Work.

RPC Remote Procedure Call.

RSA Rivest–Shamir–Adleman.

SPA Single Page Application.

UML Unified Modelling Language.

INTRODUCTION

People use the internet every day at work, at school, at home for entertainment or remote work, and almost everywhere. At first, computers were used to access the internet, the people started to share data about themselves on the internet. This data includes what they like, what they do, etc. A bit later the internet and computers started to conquer almost every field in society such as the medical field, governmental field an others. At this point, people started to host on the internet even more sensitive data than what was mentioned above, for example, medical certificates or personal IDs. These data should be protected for malicious people and those who host the data should make sure they won't lose it. There are of course developed techniques to protect the data, but it is sometimes no enough reliable.

This paper describes a system that can be used to protect data with a higher level of security. First of all, it is needed to have defined the functional and non-functional requirements for such a system. By looking at the requirements, the most appropriate technologies and techniques were defined that can be used to implement the system. The functional requirements are a few and very simple and those are to be able to write some data and to be able to retrieve back what one stored earlier. While the non-functional requirements are the most important for the project. They are what makes this project special.

By looking at the non-functional requirements and after researching on what technologies can be used, it has been determined that to always assure availability a good technology to use is IPFS because it is a protocol that allows storing data in a decentralized way, meaning that if a server falls down, the other servers (also knows as nodes) will recover the data. To assure data encryption and transparency it is possible to use Smart Contracts running on Ethereum blockchian. The Smart Contract will store the information about who owns what.

The first chapter is the domain analysis. In this chapter the problem is analysed and defined. Also in this chapter the scope and solution are being explained. Along all these, in the first chapter is explained why the user would have to pay 0.36$ to store a document on this system. The second chapter tells the functional and non-functional requirements of the system and also show the design of it using explained diagrams. As the project focuses more on showing a new approach to protect the data online, there are only a few simple functional requirements, while the non-functional requirements are the most important aspects of the project being actually the backbone definition for what should be achieved. The diagrams described how the selected technologies can be put together to implement that system that fulfils the defined requirements. The third chapter is the actual implementation of the system with all of its code examples and explanations.

Bibliography

1. *51% attack. [online] [accessed 10.10.2022].*
   Available: https://en.bitcoinwiki.org/wiki/51%25_attack.

2. *Blockchain [online] [accessed 20.10.2022].*
   Available: https://en.wikipedia.org/wiki/Blockchain.

3. *Cryptocurrency [online] [accessed 27.10.2022].*
   Available: https://en.wikipedia.org/wiki/Cryptocurrency.

4. *Facebook–Cambridge Analytica data scandal [online] [accessed 10.10.2022].*
   Available: https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal.

5. *Filecoin [online] [accessed 15.10.2022].*
   Available: https://en.wikipedia.org/wiki/Filecoin.

6. *InterPlanetary File System [online] [accessed 27.10.2022].*
   Available: https://en.wikipedia.org/wiki/InterPlanetary_File_System.

7. *IPFS Pinning Services [online] [accessed 12.11.2022].*
   Available: https://sourceforge.net/software/ipfs-pinning/.

8. *IPFS: A Complete Analysis of The Distributed Web [online] [accessed 7.10.2022].*
   Avaialble: https://medium.com/zkcapital/ipfs-the-distributed-web-e21a5496d32d.

9. *Non-fungible token [online] [accessed 05.11.2022].*
   Available: https://en.wikipedia.org/wiki/Non-fungible_token.

10. *Pin files using IPFS [online] [accessed 12.11.2022].*
    Available: https://docs.ipfs.tech/how-to/pin-files/#three-kinds-of-pins.

11. *Proof of Stake [online] [accessed 15.11.2022].*
    Available: https://en.wikipedia.org/wiki/Proof_of_stake.

12. *Proof of work [online] [accessed 15.10.2022].*
    Available: https://en.wikipedia.org/wiki/Proof_of_work.

13. *Public-key cryptography [online] [accessed 18.11.2022].*
    Avaialble: https://en.wikipedia.org/wiki/Public-key_cryptography.

14. *Satoshi Nakamoto [online] [accessed 10.10.2022].*
    Available: https://en.wikipedia.org/wiki/Satoshi_Nakamoto.

15. *Smart Contracts [online] [accessed 11.10.2022].*
Available: `https://en.wikipedia.org/wiki/Smart_contract`.

16. *Solidity programming language [online] [accessed 07.11.2022].*
Avaialble: `https://docs.soliditylang.org/en/v0.8.17/`.

17. *Sybil Attack [online] [accessed 05.11.2022].*
Available: `https://www.imperva.com/learn/application-security/sybil-attack`.

18. *Symmetric vs. Asymmetric Encryption [online] [accessed 18.11.2022].*
Available: `https://www.ssl2buy.com/wiki/symmetric-vs-asymmetric-encryption-what-are-differences`.

19. *Symmetric-key algorithm [online] [accessed 18.11.2022].*
Avaialble: `https://en.wikipedia.org/wiki/Symmetric-key_algorithm`.

20. *The Complete Open-Source and Business Software Platform [online] [accessed 12.11.2022].*
Avaialble: `https://sourceforge.net/`.