SEMINAR AND PROGRESS REPORT

MASTERS OF COMPUTER APPLICATIONS (SOFTWARE ENGINEERING)

Submitted To:-

Dr. Anju Saha
USICT, GGSIP University

Submitted By:-

Vikramaditya Bhatnagar
Roll No.:- 04816404518
MCA(SE) Sem. VI

BASIC INFORMATION ABOUT PROJECT

Project Name: Nytecrawler

Domain: Tracking-Proof File Sharing System

Technologies Used:

a. Blockchain

b. Ethereum

c. Solidity

d. IPFS

e. Metamask

BLOCKCHAIN

Wikipedia defines Blockchain as “A blockchain is a growing list of records, called blocks, that are linked together using cryptography, where each block contains a cryptographic hash of the previous block, a timestamp, and transaction data”, or alternatively as a “decentralized, distributed, and oftentimes public, digital ledger consisting of records called blocks that is used to record transactions across many computers so that any involved block cannot be altered retroactively, without the alteration of all subsequent blocks” **[https://en.wikipedia.org/wiki/Blockchain]**.

The concept of blockchain of invented by a person (or a group of persons) whose identity remains anonymous to this day, known as Satoshi Nakamoto. In 2008, he brought his vision of developing a currency that cannot be controlled by governments or banks to life; a version of currency that people are truly in control of. There are countless drawbacks of traditional currency in the real world. Some possible scenarios that can arise in the real world are –

1. Minting and printing of new currency coins and notes as per the will of a small group of people. If someone with the authority to make the decision involved wishes so, they can instruct the Reserve Bank of India to print more currency notes to be put in circulation. This results in a fall in the value of the currency, since for the same overall value a larger number of notes exist.
2. Counterfeit notes that can be put in circulation. If never identified, a counterfeit note is basically the same as an authentic note.
3. Black money

According to a white paper that Satoshi Nakamoto published on the mailing list MetzDowd.com **[Nakamoto, Satoshi (24 May 2009). "Bitcoin: A Peer-to-Peer Electronic Cash System"]**, he wished to create a form of digital gold; a resource that was limited and would always stay limited (i.e.

no new tokens of that resource could ever be manufactured) so that it would always be coveted. It also had to be something that people of influence could not control or manipulate, i.e. it had to treat every person on this planet as an equal. This led to the creation of Bitcoin, which was the world's first decentralized digital currency. It made use of the first blockchain ever created to serve as a public ledger (or a public record or every transaction that has ever happened over the Bitcoin network).

Fast forward to 2021: we have come a long way from the early days of Bitcoin. Today, the global cryptocurrency industry has a total market cap (i.e. the total amount of money invested in it) of 1.76 Trillion dollars, according to the latest June 2021 report from CoinMarketCap, the world's biggest price-tracking website by web traffic for cryptoassets and cryptocurrencies, all of which are implemented on some or the other blockchain network. CoinMarketCap also highlights in the same report that the top 5 cryptocurrencies in the world by market cap are **[https://coinmarketcap.com/about/]**:

1. Bitcoin / BTC: Market Cap of $730.5 Billion.

2. Ethereum / ETH: Market Cap of $330.1 Billion.

3. Binance Coin / BNB: Market Cap of $65.2 Billion.

4. Tether / USDT: Market Cap of $61.7 Billion.

5. Cardano / ADA: Market Cap of $58.8 Billion.

While the slots for ranks 3, 4, and 5 keep changing, the first two slots are perennially occupied in that order by the 2 titans of the cryptocurrency world: Bitcoin and Ethereum.

NEED AND JUSTIFICATION FOR USAGE

1. The Blockchain implementation is required since the feature of NFT's can only be implemented with a Blockchain. NFT's are a core part of this project, and this is the technology required.

2. Security – Blockchain is extremely secure in nature due to the way it requires multiple nodes to maintain the exact same data (all differing data is considered invalid).

3. Future Scope – Blockchain is already seeing mainstream implementation in the fields of Finance, Space Exploration, Supply Chain and Logistics Monitoring, etc., and this is in spite of it still being a bleeding-edge technology.

ETHEREUM

According to their own website **[ethereum.org]**, Ethereum define themselves as a “decentralized, open-source blockchain with smart contract functionality”.

Ethereum serves as a cryptocurrency as well as a platform for development of new cryptocurrencies and blockchain-based Decentralized Applications (also called DApps). It has attracted interested from investors and developers alike due to this reason and as a result, it is the largest blockchain network in the world today.

While the commercial (and the main) variant of Ethereum is called MAINNET, many test variants of it also exist which have been created for development and testing purposes and can be used free of cost. Some of these Ether variants are KOVAN, ROPSTEN, RINKEBY, and GOERLI. There is no conceivable difference in terms of usage and efficiency between any of these variants in terms of architecture; however, ROPSTEN has the most active faucet facility out of all these options (faucets are websites from which coins of this ETH variant can be obtained for free). Thus, the Ether test network being used in this project is ROPSTEN Ether **[G Wood - Ethereum project yellow paper, 2014]**.

Each one of these Test Ether platforms do exactly the same thing as MAINNET Ethereum, except they are free. So, just like MAINNET, the same rules apply: each one of these maintains an entire blockchain which acts as a ledger of all the transactions that have been made on this network ever since the network was created; this means that if a transaction is listed on the blockchain, it has been approved and has therefore been verified as an official transaction. No one can question the authenticity of this transaction anymore. This is an extremely secure and somewhat self-referential feature of blockchain; the blockchain itself is the verifier of the transactions.

In the Blockchain ledger, every transaction is recorded on a block. The entire ledger is the collection (or chain) of all these blocks. Think of blockchain and blocks for the purpose of a ledger as a bank passbook and its individual pages in which different transactions are chronologically written.

However, more than what meets the eye happens behind the scenes from the time you hit ‘compile’ on your IDE to the time that transaction appears on the blockchain. Every transaction has the following timeline –

1. The minute you initiate a transaction, a transaction request is sent to the back-end. Only when this request has been verified will the transaction be considered authentic.

2. In order to verify a transaction request, a certain cryptographic puzzle must be solved. This is where the name ‘cryptocurrency’ is derived from. Each cryptographic puzzle is of varying difficulty. In general, the rule is: the more lengthy and complex the transaction being tried to approve is, the more complex the puzzle will be. In the context of this project, this means that larger files will be delegated a more complex cryptographic puzzle than smaller files.

3. The solving of this puzzle is an automated process. Users can choose to donate processing power to Ethereum so that these puzzles may be solved. The more processing power you choose to donate, the higher complexity of problems you are able to solve. This is how blockchain works in a decentralized manner; it gets all its processing power from its users. In order to become one of these users who can donate processing power, you must first download the entire blockchain (i.e., *the record of literally every transaction that has occurred on this network since the beginning of time*) onto your computer. **[Ujan Mukhopadhyay, Anthony Skjellum et al: "A brief survey of Cryptocurrency systems", Dec. 2016**]. Next, be sure to have an extremely powerful computer yourself, with more-than-capable hardware like GPU's and multi-core CPU's with high clock speeds. Most regular out-of-the-box computers and laptops these days are too weak to provide the processing power necessary to be of any real use to the Ethereum network (due to its sheer size); people therefore buy dozens of expensive high-capacity GPU's just to be able to solve a few problems every day. Once you have completed all these steps, *your computer can be counted as a node of the blockchain*.

4. While this entire process of becoming a node might seem tedious, it has one huge payoff; the tantalizing promise of mining rewards. Mining in cryptocurrency is basically the process of generating new tokens (or coins) of this cryptocurrency. A simple analogy goes like this; whatever coins exchange hands via trading sites or direct wallet-to-wallet transactions are merely the equivalent of money being paid from one person to another; the mining of cryptocurrency is equivalent to the printing of currency notes (i.e., generation of new tokens). This means that you can obtain new coins of a cryptocurrency in two manners –

a. You buy it from someone who already possesses a certain quantity of said coin. This shall be the end-user way of obtaining coins, where you don't bother to study or understand the technology and end up paying a price for convenience.

b. Alternatively, you could become a node and start mining coins. While the initial investment (i.e., hardware cost) for this is extremely high, it is a one-time investment after which you basically obtain coins for free. Given enough time, you can easily make a profit through this method. *That is the single biggest motivation to start mining.* This has also had many downsides in the real world; for starters, this method of coin generation is known as PoW or Proof of Work. There have been many environmental concerns over the growing popularity of cryptocurrency, and for good reason; all this processing power involves a lot of electricity. This contributes massively to global warming. There has been a little bit of a decline in the global interest in blockchain technologies and cryptocurrency since the time the synopsis for this Major Project was penned, and that has been in large part due to heavyweight cryptocurrency advocates like Elon Musk reconsidering their stance, and ultimately deciding to dial back the enthusiasm for this bleeding-edge technology until it becomes a little more eco-friendly
[https://twitter.com/elonmusk/status/1392602041025843203?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1392602041025843203%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.bbc.com%2Fnews%2Fbusiness-57096305].

The alternative to PoW is PoS or Proof of Stake, which determines your capability to solve complex puzzles not by your processing power contributions, but by the amount of that cryptocurrency you own (i.e., your stake). For instance, someone who owns 3 Ethereum would be able to solve much more complex problems (and therefore mine more coins) than someone who only possesses 0.5 ETH. Unfortunately, most of the big names in blockchain

(including Bitcoin and Ethereum) work on a PoW model, although according to the 'Consensus Mechanism' page of the Official Ethereum Documentation, Ethereum has as of April 16, 2021 already commenced the process of shifting to a PoS model **[https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/#:~:text=Ethereum%20is%20moving%20to%20a,of%2Dwork%20(PoW)]**. This new format of Ethereum shall be known as Ethereum 2.0.

Another downside of the popularity of mining, one that might probably be seen as ludicrous by my teachers but as heart-breaking by video game enthusiasts such as myself the world over, is the rise in GPU costs because of their increased demand. GPU's or Graphical Processing Units, also called Graphic Cards, are an essential part of gaming computers and laptops as they allow us to play our favourite games in a hyper-realistic visual setting. However, rising GPU costs prohibit gamers from upgrading their personal machines so that they might enjoy the latest releases on their PC's.

The decentralized nature of blockchain ensures that no one user has complete control over the verification of transactions. Every user can verify a transaction exactly once. If a particular node has even a single transaction that does not match with the majority of the network (i.e., does not correspond with the data on the other nodes), that node is considered corrupted and is no longer a part of the network. This ensures that people cannot make fraudulent or secretive transactions by hiding them from the public ledger.

This security feature can also be interpreted as a chink in the armour that is Blockchain security; the above paragraph also means that if a majority of nodes implement a certain fraudulent change at exactly the same time on their respective nodes, it will get approved (even if the change itself is not authentic). This is known as a 51% attack, and is almost impossible to conduct on a large blockchain.

According to Swati Goyal of FxEmpire.com **[https://www.fxempire.com/education/article/51-attack-explained-the-attack-on-a-blockchain-513887#:~:text=51%25%20attacks%20on%20Bitcoin%20blockchain,miners%20all%20over%20the%20world.&text=Bitcoin%20blockchain%20has%20never%20suffered,which%20is%20hard%20to%20compromise.]**, the fact that even the most powerful computers in the world cannot compete against a pool of millions of other computers makes it extremely hard to perform such an attack. The same article says that such attacks are common in smaller blockchains with proof of work system as less computational power is required in this case. The Bitcoin blockchain has never suffered a 51% attack in part because it boasts of an active hashing power which is hard to compromise. Fortunately, it is safe to say that Ethereum – which is the largest Blockchain network in the world – is practically never going to fall prey to such an attack.

Many users worry if a PoS model would make 51% attacks easier, since now the amount of power or influence that one holds within a blockchain network is actually quantifiable and depends on the amount of cryptocurrency you own. This also conflicts with Satoshi Nakamoto's vision of an electronic cash system where all users are treated as equal **[De Filippi, Primavera, Bitcoin: A Regulatory Nightmare to a Libertarian Dream (May 14, 2014)].**

NEED AND JUSTIFICATION FOR USAGE

1. Ethereum is the largest blockchain network in the world. It is clearly extremely robust and the number one choice for building decentralized apps of this nature. It provides full end-to-end encryption and can also practically never fall prey to a 51% attack.

2. NFT's are a native concept within Ethereum; NFT's are defined by the Official Standard ERC-721 **[https://eips.ethereum.org/EIPS/eip-721]**

3. Smart Contracts are a concept that are only native to Ethereum **[https://docs.soliditylang.org/en/v0.8.4/]**

4. Metamask: Ethereum provides support for popular wallet service provider Metamask which is a key dependency for the execution of this project. This is discussed in detail in the '**METAMASK**' section of this document.

SOLIDITY

According to the Official Solidity Documentation **[https://docs.soliditylang.org/en/v0.8.4/]**, Solidity is an object-oriented programming language used for implementing smart contracts on various blockchain platforms (most notably, Ethereum). It is a curly-bracket language and is influenced by C++, Python and JavaScript. It is designed to target the Ethereum Virtual Machine (EVM). Solidity is statically typed, supports inheritance, libraries and complex user-defined types among other features. The blockchain is implemented in this project in the form of a cryptocurrency.

Solidity shall be used to write the smart contract that shall determine the validity of the trade the unique token NOTRAC over the blockchain. Smart contracts are simply programs stored on a blockchain that run when predetermined conditions are met. They typically are used to automate the execution of an agreement so that all participants can be immediately certain of the outcome, without any intermediary's involvement or time loss. They can also automate a workflow, triggering the next action when conditions are met. As a result, there are many different types of applications that can be created with the help of Smart Contracts written in Solidity, such as contracts for voting, crowdfunding, blind auctions, and multi-signature wallets.
[https://docs.soliditylang.org/en/v0.8.6/internals/optimizer.html]

On any Ethereum Network (MAINNET / ROPSTEN / KOVAN / RINKEBY / etc.), we have the concept of GAS fees; fees which are charged for carrying out a transaction **[https://itnext.io/build-a-simple-ethereum-interplanetary-file-system-ipfs-react-js-dapp-23ff4914ce4e]**. Therefore, to do even a single transaction, you need some ETH in your account (so you can pay GAS fees). The purpose of GAS is twofold; to limit the amount of work that is needed to execute the transaction and to pay for this execution at the same time.

All Smart Contracts in Ethereum are run on a runtime environment called the Ethereum Virtual Machine or EVM. It is completely isolated, which means that code running inside the EVM has no access to network, filesystem or other processes. As a result, if a smart contract (for whatever reason) crashes, the integrity of users' data, ETH holdings, and the overall Blockchain ledger will never be compromised. An instance of EVM keeps running in the background of every account created on the ETH network. It shall be this instance of EVM that shall be used for storing and subsequently running the Smart Contract for that particular account.

Ethereum's smart contracts are written in high-level programming languages and then compiled down to EVM bytecode and deployed to the Ethereum blockchain.

The Ethereum Virtual Machine has three areas where it can store data – storage, memory, and the stack.

1. STORAGE: A key-value storage space that maps 256-bit words to 256-bit words. One should avoid storing the Smart Contract in Storage since GAS prices for storing anything here are very high. Also, a contract has certain restrictions in Storage; it can neither read nor write to any storage apart from its own.
2. MEMORY: A Smart Contract obtains a freshly cleared instance of Memory for each message call, i.e. it is much more abundantly available than Storage. However, since it is scalable, the cost of creating it must be recovered by the Ethereum Network; as a result, while the initial cost of Memory is low, it scales quadratically and the GAS price can increase very quickly.
3. STACK: While the Stack is not scalable like Memory (i.e. it is of a fixed size and you must work within that), its fixed size itself is very large and therefore it is quite sufficient for the purpose of storage. Also, since it doesn't have to be scaled up GAS prices are significantly low. For all these reasons, we shall be running our Smart Contract on the STACK.

Since we store our Smart Contract in the Stack, the implementation of lines of code is very similar to implementation of code stored in a stack abstract data structure; code is popped from the stack line by line, and then the result is pushed into the stack. This result is then worked upon by the next line of code, and so on until the stack becomes empty.

NEED AND JUSTIFICATION FOR USAGE

1. While there are many high-level programming languages that can be used to write Smart Contracts in Ethereum (such as Serpent, Yul, LLL, and Mutan), Solidity is primarily used. It is clearly the most preferred choice for this purpose. This leads to many other advantages, such as its being supported by more IDE's, more resources such as tutorials to help with development, etc.

2. Smart Contracts have 2 main benefits:

- They're immutable, which means a smart contract can never be changed and no one can tamper with or break a contract.
- They're distributed, which means that the outcome of the contract is validated by everyone in the network, just like any transaction on a blockchain. Distribution makes it impossible for an attacker to force control to release funds, as all other participants would detect such an attempt and mark it as invalid.

These make them the perfect choice for specifying rules which shall apply to transactions.

3. **REMIX:** Remix is web-based compiler-cum-front end application that extends first-hand support to the Solidity programming language. Since it can be accessed on the browser, it is very convenient to use from a development point of view. Additionally, it can be connected to your Metamask wallet and can therefore be used to directly make transactions through Metamask.

IPFS

The Ethereum Problem: It is extremely expensive to store a lot of data on the Ethereum blockchain. On any variant of the Ethereum blockchain network, we have the concept of GAS fees; fees which are charged for carrying out a transaction. Therefore, to do even a single transaction, you need some ETH in your account (so you can pay GAS fees). This is also why we first add some ROPSTEN ETH to our wallet from the faucet. **[https://itnext.io/build-a-simple-ethereum-interplanetary-file-system-ipfs-react-js-dapp-23ff4914ce4e]**

According to Ethereum's yellow paper, it costs approximately 20,000 GAS to store 256 bits or 8 bytes (=1 word according to the Ethereum architecture), based on 28/02/2018 GAS prices of 4 gwei / GAS (NOTE: Current GAS prices are 13.8 gwei / GAS for standard transactions and 20.1 gwei / GAS for fast transactions). 'Gwei' basically stands for 'Giga wei', which is a denomination within the Ethereum architecture; 1 gwei = 10^(-9) ETH.

20,000 gas per Transaction of 8 bytes x 4 gwei / gas = 80,000 gwei for 8 bytes

80,000 gwei for 8 bytes. x 1000bytes/8 = 10,000,000 gwei/kB = .01 Ether

.01 Ether/kB x 1000kB = 10 Ether to store a 1Mb at $860 / Ether = $8600.00

It would cost therefore a whopping $8.6 million to store a 1 GB file on the Ethereum blockchain.

IPFS: Since it is obviously quite expensive to store data on the Ethereum blockchain, it makes much more sense to instead store the data on a decentralized file system and then share the link to the file on this file system (or the address of the file on the file system) over the Ethereum blockchain **[Zibin Zheng et al: "Blockchain challenges and opportunities: a survey", October 2018]**. For this purpose, IPFS or the Inter-Planetary File System has been used.

NEED AND JUSTIFICATION FOR USAGE

1. IPFS Address Hashes are of a fixed type and size – they are always strings with a length of 46 characters. When the header https://ipfs.io/ipfs/ is attached, this makes it a string with a total of 67 characters in it. This is not a lot of data, and therefore Gas prices for storing this data on the blockchain (and subsequently transferring it) shall also be low.

2. The file system used must be decentralized. Since Blockchain is a decentralized technology and we intend to share the link to our desired file over this decentralized technology, we are assuming that our file is stored in such a manner that it can be accessed from any computer in the entire world (i.e., it is also stored on a decentralized platform).

3. Added bonus: Unless your file is picked up by another node (i.e., you share it) or you pin it, IPFS will eventually garbage collect your file. Thus, if somehow all the other layers of security of Nytecrawler are somehow compromised and the IPFS storage address is discovered, there is a possibility that the file itself has been garbage collected by IPFS and no longer exists on the system

(possible if file has been lying around on IPFS without being shared or pinned for a relatively long period of time).

METAMASK

Whenever any cryptocurrency or token that exists on the blockchain is minted or traded, it gets associated with a certain account (which is the account of the person who currently possesses it). Since these accounts were made popular with the rise of cryptocurrencies, they are called ‘software wallets’, ‘e-wallets’, or simply ‘wallets’.

As discussed above, the Blockchain implementation is required since the feature of NFT’s can only be implemented with a Blockchain. On a blockchain, every user must create a wallet. The wallet has 2 keys – a private key (which must never be disclosed) and a public key (which acts as that wallet’s uniquely identifiable address). For the purpose of this application, 2 separate wallets have been created on 2 separate machines using the popular wallet service provider Metamask.

One might wonder how Metamask is considered a centralized technology **[https://docs.metamask.io/guide/ethereum-provider.html]**, if it serves as a wallet for tokens residing on a decentralized network. The answer is simple; the database of Metamask which is maintained by its developers and admins follows the traditional database model. This means that if somehow a data leak or hack takes place at Metamask, the account credentials of all its users would be compromised. While Metamask itself takes great care to prevent credential compromising by enabling multi-factor authentication, Nytecrawler itself has an inbuilt 6 layers of security.

NEED AND JUSTIFICATION FOR USAGE

1. Metamask is necessary when using DApps (Decentralized Applications – basically the front end of a blockchain application). There are multiple software wallets in the world, but very few provide connectivity with DApps, and Metamask is the most popular one of them all.

2. It is available as a browser extension, and is therefore easy to work with (helpful for testing prototypes on the development machine with short development cycles).

3. Provides first-hand support to both iOS and Android.

REFERENCES

https://en.wikipedia.org/wiki/Blockchain

https://en.wikipedia.org/wiki/Satoshi_Nakamoto#:~:text=Satoshi%20Nakamoto%20is%20the%20name,devised%20the%20first%20blockchain%20database.

https://coinmarketcap.com/about/

https://docs.metamask.io/guide/ethereum-provider.html

ethereum.org

https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/#:~:text=Ethereum%20is%20moving%20to%20a,of%2Dwork%20(PoW).

https://eips.ethereum.org/EIPS/eip-721

https://www.fxempire.com/education/article/51-attack-explained-the-attack-on-a-blockchain-513887#:~:text=51%25%20attacks%20on%20Bitcoin%20blockchain,miners%20all%20over%20the%20world.&text=Bitcoin%20blockchain%20has%20never%20suffered,which%20is%20hard%20to%20compromise.

De Filippi, Primavera, Bitcoin: A Regulatory Nightmare to a Libertarian Dream (May 14, 2014)

https://twitter.com/elonmusk/status/1392602041025843203?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1392602041025843203%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.bbc.com%2Fnews%2Fbusiness-57096305

Ujan Mukhopadhyay, Anthony Skjellum et al: “A brief survey of Cryptocurrency systems”, Dec. 2016

G Wood - Ethereum project yellow paper, 2014

Zibin Zheng et al: “Blockchain challenges and opportunities: a survey”, October 2018

Nakamoto, Satoshi (24 May 2009). "Bitcoin: A Peer-to-Peer Electronic Cash System"]

N Kuzmin, K Ignatiev, D Grafov - Information Science and Applications, 2020 – Springer

https://docs.soliditylang.org/en/v0.8.6/internals/optimizer.html

PROOF OF APPROVAL