A

Seminar Report

Securing IOT devices using Blockchain

Submitted in partial fulfillment of the requirements for the Award of the Degree

of

Master of Computer Applications

of

APJ Abdul Kalam Technological University

Submitted by

RAFSAL RAHIM

RegNo: TVE16MCA41

Department of Computer Applications

COLLEGE OF ENGINEERING TRIVANDRUM

AUGUST 2018

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING TRIVANDRUM

CERTIFICATE

Certified that this Seminar report entitled, ***“Securing IOT devices using Blockchain ”*** *is the paper presented by* ***“ Rafsal Rahim ”(Reg No: TVE16MCA14)*** *in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications of APJ Abdul Kalam Technological University during the year 2018.*

Prof. Baby Syla L.

Co-ordinator

Prof. Jose T Joseph.

Head of the Department

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this seminar. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my seminar.

I would like to thank **Dr.Hari V.S**, Principal, College of Engineering Trivandrum, who helped me during the entire process of work.

I am extremely grateful to **Prof.Jose T Joseph**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I would like to thank my coordinator, **Prof. Baby Syla**, Dept of Computer Applications, who motivated me throughout the work of my seminar.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this seminar. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Rafsal Rahim

Abstract

Internet of Things (IoT) plays an important role in the development of various fields. The increasing scale and scope of applications make a great demand of IoT data exchange in recent years. Meanwhile, a number of IoT data exchange platforms which dedicated to connecting various and distributed data sources are emerging. In such a platform, service providers can search and exchange the data sets that they need. However, the centralized infrastructure cannot provide enough trust as the third-party intermediaries for data exchange. As a result, most platforms unable to satisfy the complex requirements due to few institutions and individuals are willing to share their IoT data sets in such an untrustworthy environment. This paper proposes a decentralized solution based on the blockchain for IoT data trusted exchange. Specifically, In this paper, the basic principles of blockchain and corresponding key technologies are expounded through in-depth analysis of three main reliable requirements in IoT data exchange. Besides, this paper provides an architecture of above solution and detailed design of its main trust component. Finally, it realizes a prototype by using Ethereum blockchain and smart contracts and presents its auditable, transparent, decentralized features visually.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past decade, benefiting from the rapid development of wireless communication technology, sensing technology and the improvement of big data analysis capacity [1], the Internet of Things (IoT) is growing with incredible speed in most areas, especially in healthcare [2], smart city [3] and autonomous vehicles [4]. As the essential elements of the IoT world, data collected from various devices can be applied in a wide range of areas after being analyzed and processed. Combining with advanced technologies such as big data and artificial intelligence [5], IoT service based on data not only reduces the cost of industry and agriculture and makes the devices around people more intelligent, but also repeatedly optimizes the IoT ecosystem (security, equipment management and the standardization [6]) itself. However, due to the limited high maintenance and management cost [7], restricted collecting scope in privacy data, data exchange between individual and organizations become irreversible trend when they realize that connection is more important than possession [8]. Therefore, a lot of data exchange or sharing platform hava emerged in the past years. Such as crawdad for archiving wireless data [9], data science central provided industrys online resources for big data practitioners, data.gov as the home of the US governments open data platform, the development of digital coast met the unique needs of the coastal management community [10,11]. Most of them concentrated on data of the same kind or a specific field and led by government or the unions of large institutions. However, the data sets on such centralized platform cannot meet the publics diversified demands, largely due to they cannot provide enough trust to guarantee the transparency, auditable, immutable in data exchange process.

It has been watched by researchers for a long time, as the issue of trust kills the enthusiasm to share data actively and seriously hampers the development of the data industry [12]. Although current platforms integrate a lot of confidentiality mechanisms (access control, authorization privacy) and propose some trust model for data sharing in IoT, they are not break away from the third party [13].

In order to guarantee IoT data exchange in a completely trusted, transparent environment, we propose a decentralized solution based-on blockchain. As the core technology inBitcoin, blockchain soon came to be widely attracted attention and application for its trust property[14].

The core advantage of blockchain is decentralization. It consists of data encryption, timestamp, distributed consensus algorithm, economic incentive mechanism and other technology. It is applicable to the point-to-point transaction based on decentralized credit in distributed systems without mutual trust. Sequentially, this technology solves the prevailing high cost, low efficiency and uneasiness of data storage in current central institutions. This paper utilizes the feature of data is not tampered and completely transparent, combines with the time stamp and the transaction details in the process of storage and trading, so that it can be trusted by many parties. Its mixed encryption technology based on asymmetric encryption makes the user's privacy information secure with the public and private key as the only identifier of transaction subject. The second generation blockchain introduces intelligent contract, which makes the blockchain easier to use distributed application programming and speed up transaction speed. The Ethereum intelligent contract combined with capability-based access control method makes the data provider can completely control their own data sharing permissions quickly and efficiently and completely solve the credible issues of original system in the data.

Chapter 2

Little about the core technologies

2.1 Internet of Things(IOT)

2.1.1 what do IoT really Means?

Definition - What does Internet of Things (IoT) mean? The internet of things (IoT) is a computing concept that describes the idea of everyday physical objects being connected to the internet and being able to identify themselves to other devices. The term is closely identified with RFID as the method of communication, although it also may include other sensor technologies, wireless technologies or QR codes.

The IoT is significant because an object that can represent itself digitally becomes something greater than the object by itself. No longer does the object relate just to its user, but it is now connected to surrounding objects and database data. When many objects act in unison, they are known as having "ambient intelligence."

The internet of things is a difficult concept to define precisely. In fact, there are many different groups that have defined the term, although its initial use has been attributed to Kevin Ashton, an expert on digital innovation. Each definition shares the idea that the first version of the internet was about data created by people, while the next version is about data created by things. In 1999, Ashton said it best in this quote from an article in the RFID Journal:
"If we had computers that knew everything there was to know about things using data they gathered without any help from us we would be able to track and count everything, and greatly

reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best."

Most people think about being connected in terms of computers, tablets and smartphones. IoT describes a world where just about anything can be connected and communicate in an intelligent fashion. In other words, with the internet of things, the physical world is becoming one big information system.

2.1.2 How it works?

An IoT ecosystem consists of web-enabled smart devices that use embedded processors, sensors and communication hardware to collect, send and act on data they acquire from their environments. IoT devices share the sensor data they collect by connecting to an IoT gateway or other edge device where data is either sent to the cloud to be analyzed or analyzed locally. Sometimes, these devices communicate with other related devices and act on the information they get from one another. The devices do most of the work without human intervention, although people can interact with the devices – for instance, to set them up, give them instructions or access the data.

The connectivity, networking and communication protocols used with these web-enabled devices largely depend on the specific IoT applications deployed.

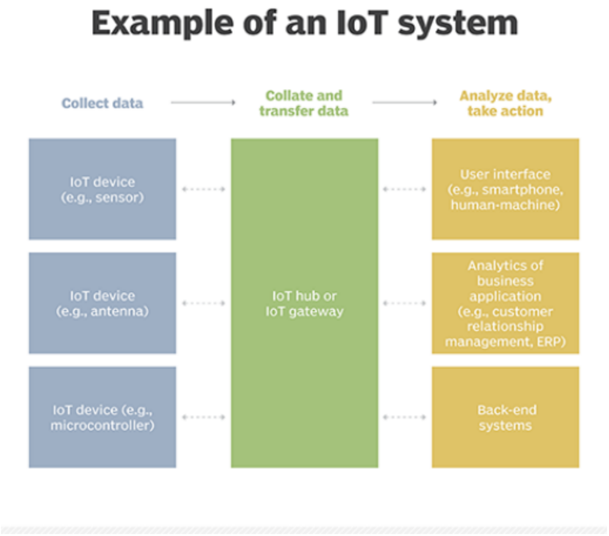

Example of an IoT system
Collect data
Collate and transfer data
Analyze data, take action
IoT device (e.g., sensor)
IoT device (e.g., antenna)
IoT device (e.g., microcontroller)
IoT hub or IoT gateway
User interface (e.g., smartphone, human-machine)
Analytics of business application (e.g., customer relationship management, ERP)
Back-end systems

Figure 2.1: IoT System

2.1.3 Consumer and enterprise IoT applications

There are numerous real-world applications of the internet of things, ranging from consumer IoT and enterprise IoT to manufacturing and industrial IoT (IIoT). IoT applications span numerous verticals, including automotive, telco, energy and more.

In the consumer segment, for example, smart homes that are equipped with smart thermostats, smart appliances and connected heating, lighting and electronic devices can be controlled remotely via computers, smartphones or other mobile devices.

Wearable devices with sensors and software can collect and analyze user data, sending messages to other technologies about the users with the aim of making users' lives easier and more comfortable. Wearable devices are also used for public safety – for example, improving first responders' response times during emergencies by providing optimized routes to a location or by tracking construction workers' or firefighters' vital signs at life-threatening sites.

In healthcare, IoT offers many benefits, including the ability to monitor patients more closely to use the data that's generated and analyze it. Hospitals often use IoT systems to complete tasks such as inventory management, for both pharmaceuticals and medical instruments.

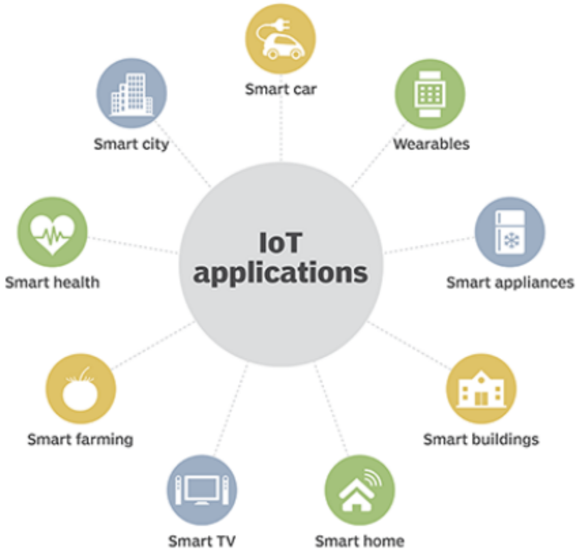


Figure 2.2: IoT Applications

Smart buildings can, for instance, reduce energy costs using sensors that detect how many occupants are in a room. The temperature can adjust automatically – for example, turning the air conditioner on if sensors detect a conference room is full or turning the heat down if everyone in the office has gone home.

In agriculture, IoT-based smart farming systems can help monitor, for instance, light, temperature, humidity and soil moisture of crop fields using connected sensors. IoT is also instrumental in automating irrigation systems.

In a smart city, IoT sensors and deployments, such as smart streetlights and smart meters, can help alleviate traffic, conserve energy, monitor and address environmental concerns, and improve sanitation.

Components of IoT:

1. **Sensors/Devices:**
   First, sensors or devices collect data from their environment. This could be as simple as a temperature reading or as complex as a full video feed.

   I use “sensors/devices, because multiple sensors can be bundled together or sensors can be part of a device that does more than just sense things. For example, your phone is a device that has multiple sensors (camera, accelerometer, GPS, etc), but your phone is not just a sensor.

   However, whether it‘s a standalone sensor or a full device, in this first step data is being collected from the environment by something.

2. **Connectivity:**
   Next, that data is sent to the cloud , but it needs a way to get there!

   The sensors/devices can be connected to the cloud through a variety of methods including: cellular, satellite, WiFi, Bluetooth, low-power wide-area networks (LPWAN), or connecting directly to the internet via ethernet.

Each option has tradeoffs between power consumption, range and bandwidth (heres a simple explanation). Choosing which connectivity option is best comes down to the specific IoT application, but they all accomplish the same task: getting data to the cloud.

3. **Data Processing:**
   Once the data gets to the cloud, software performs some kind of processing on it.

   This could be very simple, such as checking that the temperature reading is within an acceptable range. Or it could also be very complex, such as using computer vision on video to identify objects (such as intruders in your house).

   But what happens when the temperature is too high or if there is an intruder in your house? Thats where the user comes in.

4. **User Interface:**
   Next, the information is made useful to the end-user in some way. This could be via an alert to the user (email, text, notification, etc). For example, a text alert when the temperature is too high in the companys cold storage.

   Also, a user might have an interface that allows them to proactively check in on the system. For example, a user might want to check the video feeds in their house via a phone app or a web browser.

   However, its not always a one-way street. Depending on the IoT application, the user may also be able to perform an action and affect the system. For example, the user might remotely adjust the temperature in the cold storage via an app on their phone.

   And some actions are performed automatically. Rather than waiting for you to adjust the temperature, the system could do it automatically via predefined rules. And rather than just call you to alert you of an intruder, the IoT system could also automatically notify relevant authorities.

2.2 Security Issue and vulnerabilities of IoT

One of the security weakness of IoT is that it increases the number of devices behind networks firewall. As Based on the review in, ten years ago, there was a major concern about protecting computers, five years ago, the concern was about protecting smartphones. Now we have to worry about protecting our car, our home appliances, our wearables, and many other IoT devices. Computers also have security problems but with automatic and easier updates have helped alleviate this problem. But in case of IoT devices, manufacturers are pressured to get their devices in the market , there by ending up on compromising the security. Even if they may offer firmware upgrades for a time, they often stop when they focus on constructing the next device, leaving customers with slightly outdated hardware that can become a security risk.

IoT devices have security concerns, as these devices can easily get attacked by hackers, the data will be hacked and these devices may get controlled or accessed by the hackers. The point is that we have to think about what a hacker could do with a device if he can break through its security. A strong cryptographic algorithms are required to secure IoT devices and to provide secure channel. The following lists different kinds of attacks which have been observed recently and discussed in.

2.2.1 DDoS attacks

In 2016, the Mirai botnet launched one of the biggest DDoS attacks ever recorded. More than 1 terabyte per second flooded the network of Dyn, a major DNS provider, and brought down sites such as Reddit and Airnbnb. But what made this attack so special was that it was the first to be carried out with IoT devices. Nearly 150,000 compromised smart cameras, routers and other devices all enslaved into a single botnet, focused on a single target.

Manufacturers usually use a handful of default password and usernames to protect an IoT device. So there will be a few hundreds/thousands of password combinations to protect tens of millions of smart devices. All it took were a few simple lines of code, designed to test each of those default passwords. A device could be hacked and enslaved within a few seconds, so long as the user didnt change the standard login information.

2.2.2 Unsecure car apps

As Internet connected cars are coming around, it has been observed that hackers are trying to take control of the onboard software , trying to access and open the car locks. Unsecure car apps can allow malicious hackers to control ones car.

2.2.3 Insecure default login credentials

In practice, manufacturers might hide the Change password/Username options deep in the UI, out of sight for most users. If each Internet of Things device had a randomized username and password, Mirai might not have happened in the first place.

2.2.4 Poor software updates

Many Internet of Things manufacturers dont even patch or update the software that came on their devices. If a device has software vulnerability, theres little one can done to prevent an attacker from exploiting it without help from the manufacturer.

2.2.5 The communication isnt encrypted

Many IoT devices lack basic encryption to hide the data sent between the device and the central server. This can potentially expose the users personal information, if a malicious hacker can snoop in on his personal information.

Another thing that Internet of Things devices do, is that some of them ask for more permissions than they need to.One time, numerous Amazon Echo users were surprised to see their device ordering dollhouses after a TV anchor said the phrase Alexa ordered me a dollhouse.In that case, the device had permission to do a purchase all by itself. Each extra permission in an IoT device adds another vulnerability layer which can be exploited. The fewer permissions, the more secure your device is.

2.2.6 Insecure user interface

A devices user interface is usually the first thing a malicious hacker will look into for any vulnerabilities. For instance, he might try to manipulate the I forgot my password, in order to reset it or at least find out your username or email.

A properly designed device should also lock out a user from attempting to login too many times. This stops dictionary and brute force attacks that target passwords, and greatly secures your device credentials. In other cases, the password might be sent from the device to the central server in plain text, meaning it isnt encrypted.

2.2.7 Poor privacy protection

Internet connected devices are data-hungry beasts, but some of them have a greater appetite than others. The less information they have on you, the better, since it limits how much a cybercriminal can learn about you if he hacks the device. As a rule, try to look into what type of data a device will store about you. Be critical of those that harvest data they dont need, such as coffee machines storing users location information.

2.3 Blockchain

2.3.1 Structure of a block

A blockchain, is a growing list of records, called blocks, which are linked using cryptography. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a merkle tree root hash). A block is a container data structure, which brings together transactions for inclusion in the public ledger, known as the blockchain. The block is made up of a header; containing metadata, followed by a long list of transactions. A block can be identified in two ways, either by referencing the block hash, or through referencing the block height. The block header consists of three sets of block metadata. Metadata is data that provides information about other data. Firstly, there is a reference to a previous block hash, which connects this block to the previous block, lying in the blockchain. The second set of metadata relates to the mining competition; namely the difficulty, timestamp and nonce. Lastly, the third piece of metadata is the Merkle Tree root; a data structure used to summarize all the transactions in the block in an efficient manner.

Block headers can be regarded as an example of a dynamic membership multi-party signature (DMSS). A DMSS is a digital signature formed by a set of signers which has no fixed size (Back, Corallo, Dashjr, Friedenbach, 2014).Bitcoins block headers are DMSS because their proof of work has the property that anyone can contribute without undergoing an enrolment process. Furthermore, contribution is weighted by proportional computational power rather than one threshold signature contribution per party (Back, Corallo, Dashjr, Friedenbach, 2014).

This allows anonymous membership without risk of a Sybil attack. A Sybil attack is when one party joins many times and has an uneven, disproportionate input into the signature. Since the blocks are chained together, Bitcoins DMSS is cumulative. A chain of block headers is also a DMSS on its first block, with computational strength equivalent to the sum of the computational strengths of the composing DMSS . Therefore, the key innovation in Blockchain is a signature of computational power, rather than the typical signature of knowledge.

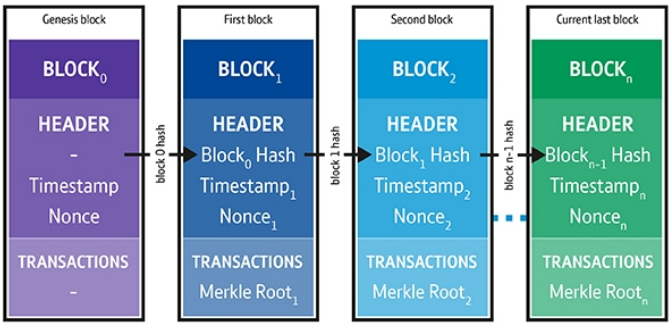


Figure 2.3: Blockchain structure

2.3.2 Block header hash and nodes

Here I have am providing an example, the block hash of the first Bitcoin block ever created will be like 000000000019d7789c085ae165831e934gf763ae46a4a6c172b3f1b60a8ce26f. The block hash identifies a block uniquely, and can be independently derived by any node simply by hashing the block header. A node is a full client. A full client is a client that owns the block chains and is sharing blocks and transactions across the blockchain network. A node is considered to be part of the blockchain infrastructure, and does not necessarily have to be a miner. Each node keeps a complete copy of a totally ordered sequence of events in the form of a blockchain . The blocks hash is computed by each node, as the block is received from the network. The block hash may be stored in a separate database table as part of the blocks metadata, to facilitate indexing and faster retrieval of blocks from disk.

2.3.3 Block height

Block height is another method to identify a block, this time through its position in the blockchain. The first block ever created is at block height 0 (zero), and in the case of Bit- coin, is the same block that was referenced by the block hash of the above block which is 000000000019d7789c085ae165831e9 .Each subsequent block added on top of that first block is one position higher in the blockchain, like boxes stacked one on top of the other.Block height does not always identify a particular singular block. It is possible for two or more blocks may have the same block height, both competing for the same position in the blockchain.

2.3.4 Genesis Block

The first block in any blockchain is termed the genesis block. If you start at any block and follow the chain backwards chronologically, you will arrive at the genesis block. The genesis block is statically encoded within the client software, that it cannot be changed. Every node can identify the genesis blocks hash and structure, the fixed time of creation, and the single transactions within. Thus every node has a secure root from which is possible to build a trusted blockchain on.

2.3.5 Proof of Work

In Proof of Work, in order for an actor to be elected as a leader and choose the next block to be added to the blockchain they have to find a solution to a particular mathematical problem. Given that the hash function used is cryptographically secure, the only way to find a solution to that problem is by bruteforce (trying all possible combinations). In other words, probabilistically speaking, the actor who will solve the aforementioned problem first the majority of the time is the one who has access to the most computing power. These actors are also called miners. It has been widely successful primarily due to its following properties:

1. It is hard to find a solution for that given problem

2. When given a solution to that problem it is easy to verify that it is correct.

Whenever a new block is mined, that miner gets rewarded with some currency (block reward, transaction fees) and thus are incentivized to keep mining. In Proof of Work, other nodes verify the validity of the block by checking that the hash of the data of the block is less than a preset number.Due to the limited supply of computational power, miners are also incentivized not to cheat. Attacking the network would cost a lot because of the high cost of hardware, energy, and potential mining profits missed.

2.3.6 Linking blocks in the blockchain

Nodes maintain a copy of the blockchain locally, starting from the genesis block. The local copy of the blockchain constantly updates as new blocks are discovered and subsequently built on the chain. As a node receives information of incoming blocks from the network, it will validate these blocks first, then link them to the existing blockchain.

The process to establish a link is as follows; a node will examine the incoming block header and look for the previous block hash. Looking at this incoming block, the node finds the previous block hash field, which contains the hash of its parent block. This hash is known to the node previously. Therefore, the node reasons that this new block is a child of the last block on the chain, and is the legitimate extension of the chain. The node adds this new block to the end of the chain, making the blockchain longer with a new height of the incoming block, now validated.

Chapter 3

Existing System Approach

In existing system data reliability, data availability, and response time are dealt with data replication strategies. However, storing replicas data over a number of nodes increases the attack surface for that particular data.For example, storing m replicas of a file in a cloud instead of one replica increases the probability of a node holding file to be chosen as attack sufferer, from 1/n to m/n where n is the total number of nodes. Existing system was not achieving proper security.

3.1 Disadvantages

- A key factor determining the throughput of a cloud that stores data is the data retrieval time.
- In large-scale systems, the problems of data reliability, data availability, and response time are dealt with data replication strategies.
- However, placing replicas data over a number of nodes increases the attack surface for that particular data.
- Affected on security and performance.

3.2 Data Fragmentation

The security of a large-scale system, such as cloud depends on the security of the system as a whole and the security of individual nodes. A successful intrusion into a single node may have severe consequences, not only for data and applications on the victim node, but also for the other nodes. The data on the victim node may be revealed fully because of the presence of the whole file. A successful intrusion may be a result of some software or administrative vulnerability. In case of homogenous systems, the same flaw can be utilized to target other nodes within the system. The success of an attack on the subsequent nodes will require less effort as compared to the effort on the first node. Comparatively, more effort is required for heterogeneous systems. However, compromising a single file will require the effort to penetrate only a single node. The amount of compromised data can be reduced by making fragments of a data file and storing them on separate nodes. A successful intrusion on a single or few nodes will only provide access to a portion of data that might not be of any significance. Moreover, if an attacker is uncertain about the locations of the fragments, the probability of finding fragments on all of the nodes is very low. Consider a cloud with M nodes and a file with z number of fragments. Let s be the number of successful intrusions on distinct nodes, such that s ¿ z

The probability that s number of victim nodes contain all of the z sites storing the file fragments (represented by $P(s, z)$)is given as:

$$P(s,z) = \frac{\begin{pmatrix} s \\ z \end{pmatrix} \begin{pmatrix} M-s \\ s-z \end{pmatrix}}{\begin{pmatrix} M \\ s \end{pmatrix}}$$

3.3 Centrality

The centrality of a node in a graph provides the measure of the relative importance of a node in the network. The node is important in the network if it:
(a) interconnects more nodes than others,
(b) can be reached easily by other nodes, or
(c) can reach other nodes easily.

The objective of improved retrieval time in replication makes the centrality measures more important. There are various centrality measures; for instance, closeness centrality, degree centrality, betweenness centrality, eccentricity centrality, and eigen- vector centrality. Here only elaborate on the closeness, betweenness, and eccentricity centralities because here using the aforesaid three centralities in this work.

3.3.1 Betweenness Centrality

The betweenness centrality of a node n is the number of the shortest paths, between other nodes, passing through n. Formally, the betweenness centrality of any node v in a network is given as:

$$C_b(v) = \sum_{a \neq v \neq b} \frac{\delta_{ab}(v)}{\delta_{ab}}$$

3.3.2 Closeness Centrality

A node is said to be closer with respect to all of the other nodes within a network, if the sum of the distances from all of the other nodes is lower than the sum of the distances of other candidate nodes from all of the other nodes. The lower the sum of distances from the other nodes, the more central is the node. Formally, the closeness centrality of a node v in a network is dened as:

$$C_c(v) = \frac{N-1}{\sum_{a \neq v} d(v,a)}$$

where N is total number of nodes in a network and d(v,a)represents the distance between node v and node a.

3.3.3 Eccentricity

The eccentricity of a node n is the maximum distance to any node from a node n. A node is more central in the network, if it is less eccentric. Formally, the eccentricity can be given as:

$$E(v_a) = max_b \; d(v_a, v_b)$$

3.4 T-Coloring

Consider a graph G=(V,E)and a set T containing non-negative integers including 0. The T-coloring is a mapping function f from the vertices of V to the set of non-negative integers, such that f(x)- f(y) is not an element of T, where(x,y) is an element of E. The mapping function f assigns a color to a vertex. In simple words, the distance between the colors of the adjacent vertices must not belong to T. Formulated by Hale, the T-coloring problem for channel assignment assigns channels to the nodes, such that the channels are separated by a distance to avoid interference.

Chapter 4

Proposed System

The proposed system called DROPS(Division and Replication of Data in Cloud for Optimal Performance and Security) jointly approaches the security and performance issues. The proposed DROPS scheme ensures that even in the case of a successful attack, no meaningful information is disclosed to the attacker. The DROPS technique doses not depend on traditional cryptographic techniques for data security. The non-cryptographic nature of the proposed scheme makes it faster to perform the required operations (placement and retrieval) on the data. It make sure a controlled replication of the file fragments, where each of the fragments is replicated only once for the purpose of improved security. A cloud storage security scheme jointly deals with the security and performance in terms of retrieval time.

4.1 Advantages

- Improve security.
- Improve performance.
- No information is revealed to the attacker. (If an attacker is uncertain about the locations of the fragments, the probability of finding fragments on all of the nodes is very low.)
- No load on single node of cloud.
- Numbers of fragments are decided according to owners choice.

Chapter 5

DROPS System Model

A cloud that consists of M nodes, each with its own storage capacity. Let Si represents the name of ith node and si denotes total storage capacity of Si Communication time between Si and Sj is the total time of all of the links within a selected path from Si to Sj represented by c(i, j). Consider N number of file fragments such that Ok denotes k -th fragment of a file while ok represents the size of k-th fragment. Pk denote the primary node that stores the primary copy of Ok, replication scheme for Ok denoted by Rk is also stored at Pk and Whenever there is an update in not as an independent document. Please do not revise any of the current designations Ok , the updated version is sent to Pk that broadcasts the updated version to all of the nodes in Rk. Let colSi store the value of assigned color to Si. The colSi can have one out of two values, namely: open color and close color. The value open color represents that the node is available for storing the file fragment. The value close color shows that the node cannot store the file fragment The set T is used to restrict the node selection to those nodes that are at hop-distances not belonging to T.

In the DROPS methodology, not to store the entire file at a single node. The DROPS methodology fragments the file and makes use of the cloud for replication. The fragments are distributed such that no node in a cloud holds more than single fragment, so that even a successful attack on the node leaks no significant information.

Symbols	Meanings
M	Total number of nodes in the cloud
N	Total number of fragments to be placed
O_k	k-th fragment of the file
o_k	Size of O_k
S^i	i-th node
s_i	Size of S^i
cen_i	Centrality measure of S^i
col_{S^i}	Color assigned to S^i
r_k^i	Number of reads for O_k from S^i
R_k^i	Aggregate read cost of r_k^i
w_k^i	Number of writes for O_k from S^i
W_k^i	Aggregate write cost of w_k^i
NN_k^i	Nearest neighbour of S^i holdingO_k
c(i,j)	Communication cost between S^i and S^j
P_k	Primary node for O_k
R_k	Replication schema for O_k
RT	Replication Time

Table 5.1: Notations and their meanings.

In the DROPS methodology, user sends the data file to cloud. The cloud manager system(a user facing server in the cloud that entertains users requests) upon receiving the file performs:

1. fragmentation

2. first cycle of nodes selection and stores one fragment over each of the selected node and

3. second cycle of nodes selection for fragments replication. The cloud manager keeps record of the fragment placement and is assumed to be a secure entity.

Chapter 6

DROPS Architecture

This system mainly consist of 3 modules:

6.1 Cloud Client

Cloud client should be Data owner or Data user.

6.1.1 Data Owner

Data owner is responsible for uploading file on cloud as well as view files uploaded by him or others. Data owner has information about the placed fragment and its replicas with their node numbers in cloud.

6.1.2 Data User

Data user is the one who is responsible for downloading files or view files uploaded by others. To download file from cloud he has to be authenticated user otherwise he will be considered as attacker.

6.2 Cloud Server

6.2.1 Fragmentation

This approach is used for fragmenting the file for security purpose at sever side. This approach runs the Fragmentation algorithm. It has file as input and produces the file fragments as output.

6.2.2 Replication

This approach creates replicas (duplicate copy) of fragments. These replicas are useful when one of fragment is corrupted by attacker then to provide file for user admin replaces its replica at that place and combine all fragments and send file to authenticated user or data owner. To make replicas of file fragments this approach runs replication algorithm which takes input as fragments and produces its replicas as output.

6.2.3 Allocation

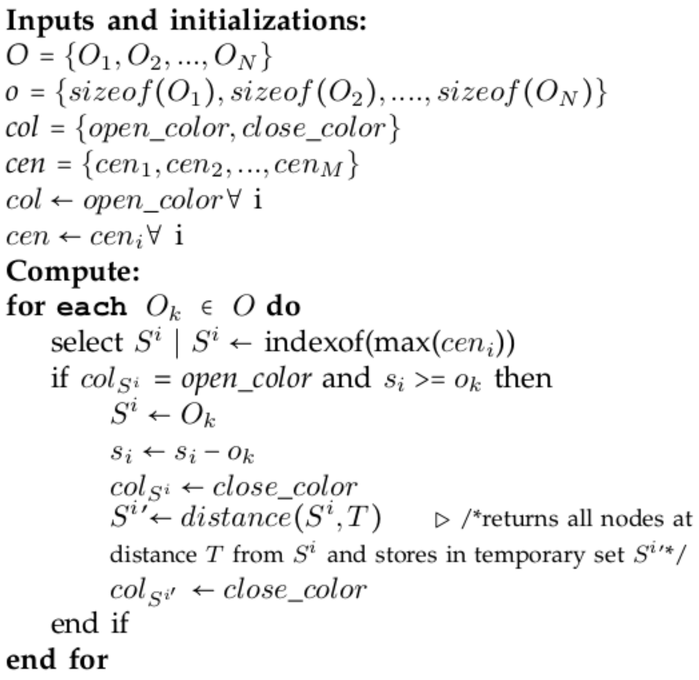
After the file is spitted and replicas are generated then allocate that fragments at cloud server for storing data. While storing or allocating that fragments, consider security issues. So the proposed method using T- Coloring Graph concept for placing fragments at different nodes on cloud server. This approach runs Fragment allocation algorithm which takes input as fragments and produces the output as fragments allocated with node numbers.

6.3 Admin

Admin is an authorized person who has rights to validate authorized data owner and user. He is also responsible for allocation of block and maintains information and authentication.

6.4 Fragment Placement

In the DROPS methodology, not to store the entire le at a single node. The DROPS methodology fragments the le and makes use of the cloud for replication. The fragments are distributed such that no node in a cloud holds more than a single fragment, so that even a successful attack on the node leaks no signicant information. The DROPS methodology uses controlled replication where each of the fragments is replicated only once in the cloud to improve the security. Although, the controlled replication does not improve the retrieval time to the level of full-scale replication, it signicantly improves the security.The fragment placement strategy is presented in the algorithm given below.



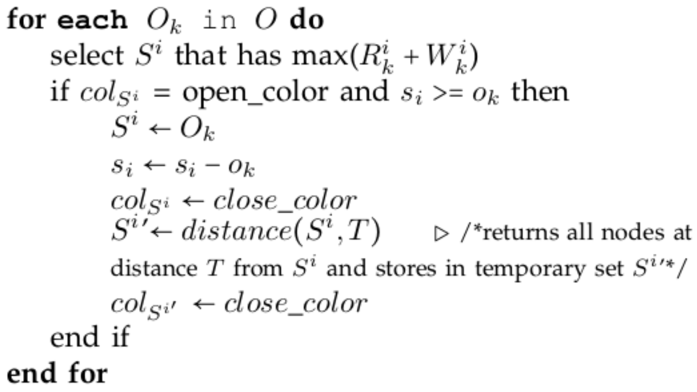
```
Inputs and initializations:
$O = \{O_1, O_2, ..., O_N\}$
$o = \{sizeof(O_1), sizeof(O_2), ...., sizeof(O_N)\}$
$col = \{open\_color, close\_color\}$
$cen = \{cen_1, cen_2, ..., cen_M\}$
$col \leftarrow open\_color \forall$ i
$cen \leftarrow cen_i \forall$ i
Compute:
for each $O_k \in O$ do
    select $S^i \mid S^i \leftarrow$ indexof(max($cen_i$))
    if $col_{S^i} = open\_color$ and $s_i >= o_k$ then
        $S^i \leftarrow O_k$
        $s_i \leftarrow s_i - o_k$
        $col_{S^i} \leftarrow close\_color$
        $S^{i\prime} \leftarrow distance(S^i, T)$    ▷ /*returns all nodes at
        distance $T$ from $S^i$ and stores in temporary set $S^{i\prime}$*/
        $col_{S^{i\prime}} \leftarrow close\_color$
    end if
end for
```

Figure 6.1: Algorithm for Fragment Placement

Initially, all of the nodes are given the open color. Once a fragment is placed on the node, all of the nodes within the neighborhood at a distance belonging to T are assigned close color. In the aforesaid process, lose some of the central nodes that may increase the retrieval time but achieve a higher security level. If somehow the intruder compromises a node and obtains a fragment, then the location of the other fragments cannot be de- termined. The attacker can only keep on guessing the location of the other fragments. However,the probability of a successful coordinated attack is extremely minute. The pro- cess is repeated until all of the fragments are placed at the nodes. The above algorithm represents the fragment placement methodology.

6.5 Fragment Replication

In addition to placing the fragments on the central nodes, this approach also per- form a controlled replication to increase the data availability, reliability, and improve data retrieval time. Place the fragment on the node that provides the decreased access cost with an objective to improve retrieval time for accessing the fragments for reconstruction of original le. While replicating the fragment, the separation of fragments as explained in the placement technique through T- coloring, is also taken care off.



```
for each $O_k$ in $O$ do
    select $S^i$ that has max($R^i_k + W^i_k$)
    if $col_{S^i}$ = open_color and $s_i >= o_k$ then
        $S^i \leftarrow O_k$
        $s_i \leftarrow s_i - o_k$
        $col_{S^i} \leftarrow close\_color$
        $S^{i\prime} \leftarrow distance(S^i, T)$    ▷ /*returns all nodes at
        distance $T$ from $S^i$ and stores in temporary set $S^{i\prime}$*/
        $col_{S^{i\prime}} \leftarrow close\_color$
    end if
end for
```

Figure 6.2: Algorithm for Fragment Replication

In case of a large number of fragments or small number of nodes, it is also possible that some of the fragments are left without being replicated because of the T-coloring. As discussed previously, T-coloring prohibits to store the fragment in neighborhood of a node storing a fragment, resulting in the elimination of a number of nodes to be used for storage. In such a case, only for the remaining fragments, the nodes that are not holding any fragment are selected for storage randomly. The replication strategy is presented in the algorithm given below.

6.6 Replication Time

Aim is to minimize the overall total network transfer time or replication time (RT) or also termed as replication cost (RC). The RT is composed of two factors: (a) time due to read requests and (b) time due to write requests.

- The total read time of Ok by Si from NNi k is denoted by Ri k and is given by:

$$R^i_k = r^i_k o_k c(i, NN^i_k)$$

- The total time due to the writing of Ok by Si ad- dressed to the Pk is represented as Wi k and is given:

$$W^i_k = w^i_k o_k (c(i, P_k) + \sum_{(j \in \neq R_k), j \neq i} c(P_k, j))$$

- The overall RT is represented by:

$$RT = \sum_{i=1}^{M} \sum_{k=1}^{N} (R^i_k + W^i_k)$$

- The storage capacity constraint states that a file fragment can only be assigned to a node, if storage capacity of the node is greater or equal to the size of fragment.
- To handle the download request from user, the cloud manager collects all the fragments from the nodes and reassemble them into a single file. Afterwards, the file is sent to the user.

6.7 Features of DROPS

- It ensures that even in case of a successful attack, no meaningful information is revealed to the attacker.
- A successful attack on a single node must not reveal the locations of other fragments within the cloud.
- The nodes storing the fragments, are separated with certain distance by means of graph T-coloring to prohibit an attacker of guessing the locations of the fragments.
- Does not rely on the traditional cryptographic techniques for the data security.
- Faster
- Higher level of security with slight performance overhead.

Chapter 7

Conclusion and future scope

The DROPS methodology is a cloud storage security scheme that jointly deals with the security and performance in terms of retrieval time. The data file was fragmented and the fragments are scattered over multiple nodes. The nodes were separated by means of T-coloring. The fragmentation and dispersal make sure that no significant information was obtainable by an antagonist in case of a successful attack. No node in the cloud, stored more than a single fragment of the same file.

The scope of this work is divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a single fragment of a particular data file that ensures that even in case of a successful attack, no meaningful information is revealed to the attacker.

Bibliography

[1] Lade Mayuri S.Lolage Pranali S. Nagare Mayuri S.Sadaphal Priyanka V. Prof.Jorwekar Y.S, *"An Study on-Cloud Security Using DROPS Technique"*, Vol. 2 Issue. 6, 2016.

[2] Priya Patni, Dr. S.N Kakarwal , *"Optimal performance and security by division and replication of data in cloud"*, Global Journal of Engineering Science and Research Management, Nov. 2015.

[3] Nidhi Jain, Archana jadhav, *"A Survey Paper on Drops: Division and Replication of Data in Cloud for Optimal Performance and Security"* , International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, Oct. 2016

[4] Mazhar Ali, K. Bilal, S. U. Khan, B. Veeravalli, K. Li, Albert Y. Zomaya and A. J. R. Neves, *"dropd: Division and Replication of Data in the Cloud for Optimal Performance and Security"*, IEEE transactions of cloud computing., vol. 13, no. 11, pp. 14111418, Nov. 2015.