

SECURE FILE SHARING AND RETRIEVAL IN CYBER PHYSICAL CLOUD SYSTEMS

A PROJECT REPORT

Submitted by

JAYARAMAN V (2016503013)

MEENAKSHI SUNDARAM S (2016503025)

YAMINI O (2016503553)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING



DEPARTMENT OF COMPUTER TECHNOLOGY

ANNA UNIVERSITY, MIT CAMPUS

CHENNAI 600 044

OCTOBER 2020

ANNA UNIVERSITY : CHENNAI 600 044

BONAFIDE CERTIFICATE

Certified that this project report “**SECURE FILE SHARING AND RETRIEVAL IN CYBER PHYSICAL CLOUD SYSTEMS**” is the bonafide work of “**JAYARAMAN V (2016503013), MEENAKSHI SUNDARAM S (2016503025) and YAMINI O (2016503553)**” who carried out this project under my supervision.



SIGNATURE

DR. R GUNASEKARAN
HEAD OF THE DEPARTMENT

Professor

Department of Computer Technology

Anna University, MIT Campus

Chromepet, Chennai - 600044



SIGNATURE

DR. PONSY R K SATHIA BHAMA
SUPERVISOR

Associate Professor

Department of Computer Technology

Anna University, MIT Campus

Chromepet, Chennai - 600044

ABSTRACT

Effective file sharing between any lightweight devices still remains a problem to be addressed. Now-a-days data storing is becoming a sensitive issue to handle due to the rise in technology which resulted in the compromise of security. Storing the raw data as such in any form results in data loss if the storage security has been compromised. Heavyweight devices can be made secure by producing heavy weight algorithms but the same situation becomes a different one to handle for lightweight devices. Hence some considerable amount of security needs to be incorporated to handle sensitive data. Thus, the proposed system provides effective measures to share the files in cyber physical cloud systems, among the owners and consumers through encryption and also provides effective multi-keyword semantic search to query the files present on the cloud storage.

ACKNOWLEDGEMENT

We thank the teaching and non-teaching members those who belong to the Computer Centre, Applied Science & Humanities of Madras Institute of Technology, Anna University for shaping our initial years of college life.

We are highly indebted to our respected Dean, **DR. T THYAGARAJAN** and to our respected Head of the Department, **DR. R GUNASEKARAN** Department of Computer Technology, MIT, Anna University for providing sufficient facilities that contributed to the success of the project.

We express our deep sense of gratitude to our guide **DR. PONSY R K SATHIA BHAMA**, Associate Professor, Department of Computer Technology, Madras Institute of Technology, Anna University for her guidance, vision and constant encouragement throughout our project.

We thank our panel members **DR. R VARALAKSHMI**, Associate Professor, **DR. V P JAYACHITRA**, Assistant Professor for their valuable comments and their different views on our project.

Finally, we thank all the teaching and non-teaching members of the Department of Computer Technology, library staff members, friends, seniors, juniors, and everyone who has bestowed us with their knowledge directly or indirectly, intentionally or unknowingly.

JAYARAMAN V (2016503013)

MEENAKSHI SUNDARAM S (2016503025)

YAMINI O (2016503553)

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iii
	LIST OF FIGURES	viii
	LIST OF ACRONYMS AND ABBREVIATIONS	ix
1.	INTRODUCTION	1
	1.1 Overview	1
	1.2 Encryption and Decryption	2
	1.3 Types of encryption	3
	1.4 Public Key encryption	3
	1.5 Symmetric Encryption	4
	1.6 File Processing	5
	1.7 NLP in files	6
	1.8 Architecture design	6
2.	LITERATURE SURVEY	8
	2.1 Overview	8
	2.2 Related Works	8
3.	PROPOSED WORK	15
	3.1 Overview	15
	3.2 Ideology	15
	3.3 User Registration	16
	3.3.1 Introduction	16

3.3.2 Process	16
3.4 Mutual Authentication	17
3.4.1 Introduction	17
3.4.2 Process	17
3.5 Encryption	18
3.5.1 Introduction	18
3.5.2 Process	19
3.5.3 Countering Disadvantages	20
3.5.4 Reason for ECC	21
3.6 File Retrieval	21
3.6.1 Introduction	21
3.6.2 Process	21
3.6.3 File Processing	22
3.6.4 Text Mining	22
3.6.5 Image Files	23
3.7 Architecture Diagram	23
3.7.1 Overview	23
3.7.2 Diagram	24
3.7.3 Workflow diagram	25
4. IMPLEMENTATION DETAILS	26
4.1 Overview	26
4.2 Mutual Authentication	26
4.2.1 Diffie-Helman	26
4.2.2 Conclusion	27

4.3	Encryption	28
4.3.1	Introduction	28
4.3.2	Why ECC? Why Not RSA	28
4.3.3	ECC	28
4.3.3.1	Assumptions	28
4.3.3.2	Point Generation	30
4.3.3.3	Key Generation	32
4.3.3.4	Encryption and Decryption	32
4.3.3.5	Montgomery Ladder	33
4.3.3.6	Comb Multiplication	34
4.4	Configuration File Generation	36
5.	EXPERIMENTAL RESULTS	38
5.1	Overview	38
5.2	Encryption and decryption	38
5.3	Configuration File Generation	39
6.	CONCLUSION AND FUTURE WORKS	41
	REFERENCES	42

LIST OF FIGURES

FIGURE NO	FIGURE TITLE	PAGE NO
3.1	Architecture diagram of secure file sharing in CPC	24
3.2	Workflow diagram for secure file sharing in CPC	25
4.1	Elliptic curve Point Generation	30
5.1	Input file for ECC encryption	38
5.2	File Encrypted by ECC	39
5.3	Input text file for Configuration file generation	39
5.4	Generated Configuration file	40

LIST OF ACRONYMS AND ABBREVIATIONS

ACRONYMS	ABBREVIATIONS
ABE	Attribute Based Encryption
CPS	Cyber Physical Systems
DH	Diffie-Helman
ECC	Elliptic Curve Cryptography
ESSI	Extended Service Set Image
MCC	Mobile Cloud Computing
IBE	Identity Based Encryption
KDC	Key Distribution Center
PAKE	Password Authenticated Key Exchange
RSA	Rivest Shamir Adleman
ROM	Random Oracle Model
SSE	Searchable symmetric Encryption
SN	Service Node

CHAPTER 1

INTRODUCTION

1.1 Overview

Information and data are the most vitally considered elements of modern age. Every process that we perform needs data, produces data, and works on data. Some data that are used in new ages consider confidential information, maybe about the user that can lead to the expulsion of the identity of the user. Hence either the process should be made secure not to work on multifarious data that acts as a threat to identity or the data feeded to the system must be in a way that even after they have been leaked they must in a secured manner such that the receptor can identify the data only if the receptor is a valid user. Since the former is more tough to implement that latter as the personalisation which is making a major turn over in the software industry has no other way rather than working with sensible data containing user identity.

Many mathematicians have worked hard to produce models that can be used as a virtual lock to data, such that even if the data is present and can be seen it will be in an uninterpretable form. These works are later termed as cryptographic algorithms. The complexity in encrypting the data remains as still while decrypting, hence the way to minimise the complexity and time to decrypt the data along with providing the same level of security in data is what the industries are moving into. Nowadays many user data have been stored in a third party server or with a trusted authority, hence providing the proper level of security mechanisms are vital part in today's technological survival and salvation of software industries.

1.2 Encryption and Decryption

As the people started relying more on the computer systems that work on algorithms to store their personal data, it became a new goal for many researchers in providing a much safer Cyber Physical Systems(CPS) that makes use of the detailed and more reliable models of mathematics and many proposed impenetrable cryptographic algorithms to satisfy the users in a permutural way.

Many detailed procedures have been documented in stealing of data and also unprecedented data loss. The way of approaching the problem through blacklisting the illegal methods will not be an advisable solution, as in earlier stages it provided a certain level of assurance, in the modern era every day new technologies have been found to work on the complex security algorithms and high level mathematical models, more than hundred new vulnerabilities have been exploited to steal data away resulting in a well planned data loss. Hence whitelisting the vulnerability will be a proper solution, i.e , pre-empting the danger and tackling it rather than relying on the lock alone the key to unlock the lock is also preprocessed such that not everyone with the key will be provided the access to the data.

These techniques have been widely adopted since the attack that has been performed in Microsoft Dropbox, a data storing platform that has been proved untrustworthy. A novel attack called heartbleed attack performed lately which exposed the vulnerable side of networks of the tech giant, pushed the researchers in finding a proper solution to the problem and keeping the user data in a safer manner. Since the data stored are to be retrieved, the focus must be in reduction of the retrieval time of the data, which leads to the introduction of a new method called Searchable encryption methods.

1.3 Types of encryption

There are two ways of encrypting a data in a safer manner, namely symmetric encryption and asymmetric encryption. Each method has their own level of advantages and disadvantages, hence the place where they have been used is a key factor in measuring its effectiveness.

1.4 Public key encryption

In 1976, Whitfield Diffie, Marty Hellmann and then graduate student Ralph Merkle who are nowadays termed as fathers of modern cryptography, came up with a breakthrough idea called Public Key Cryptography which is a most reliable form in today's security processes. Thus, in order to transmit a data secretly, it requires three major elements namely, the data to move from point A to point B, a cryptographic algorithm that scrambles the information such that it makes impossible for middleman to read, and a special element called cryptographic key which locks and unlocks the unreadable form of data in destination to an interpretable one. The method still remains a notable one but the main problem these methods are undergoing are the keys are being predicted and cracked by some highly qualified nefarious actors. Keys are nothing but typically large numbers mostly primes, which undergo a trapdoor operation in which performing encryption is easy whereas decrypting becomes a tedious one (example, multiplication of primes is easy, factoring and decrypting back to the primes are tedious ones). Over the years it was effective as the encryption was performed with smaller numbers, but the introduction of supercomputers and many quantum revolutions, the safety became compromised as the keys were cracked very easily. Hence each time a key keeps getting cracked, doubling of size is performed, expecting that the time taken to crack the key of

this size will be improbable. This has been going on for 40 years and the last number that has been cracked is seven hundred and sixty digits long.

1.5 Symmetric encryption

Although symmetric key and asymmetric key cryptography relies on keys for encryption and decryption, public key cryptography works on two kinds of keys to encrypt and decrypt data, namely public and private keys whereas symmetric key cryptography uses only one key but changes the procedure every time of decryption. Although it's easy to exploit the symmetric key framework by cracking only one key, the main advantage in symmetric key cryptography is that they are easy to build on and easy to store the key so that the programmer can focus on fortifying the process with more sophisticated algorithms. The other main problem which the symmetric encryption encountered is the key sharing process as it involves only one key for encryption and decryption, the receptor needs to know the key beforehand or the key needs to be transmitted along, thus making the system vulnerable while transmission. There are many algorithms proposed along for this kind of system namely AES(Asymmetric Encryption Standard), DES(Digital Encryption Standard), DSA(Digital Signature Algorithm). These algorithms primarily work on keys, as they build tough algorithms that perform divide and conquer methods, the keys are splitted and many operations like XOR, AND etc, have been performed (normally consists of 64 to 128 steps) to yield an encrypted format. The same processes are reversed with proper care to decrypt to the same form.

Even the public key cryptography methods like RSA, ECC etc, have their own disadvantages, the whole internet uses them as a basic plot to build a secure fortress in order to exploit their easy way of transmitting the keys. For this

purpose, a new category of algorithms originated naming themselves as key exchange algorithms.

1.6 File processing

The data or information that has been encrypted and made safe is such that we can use them for future access. Thus proper steps should be carried out while performing encryption such that no information is lost in any steps that cannot be regained. The loss of data is one of the widely encountered problems in this field, during the operation any block memory that is read is directly written instead of writing its copy leading to a bulk data loss. Thus an architecture should be proposed in which a level of preprocessing is made before encryption such that it will not only be easy to ensure the presence of data but also provides a better way of searching the file among large piles of data. The preprocessing includes creating a tag file for every encrypted file which contains information about what the file contains and some gist of the file.

The next obstacle in this kind of approach is securing the configuration file. Since they also cannot be encrypted, they must be made object readable such that no other third party can access the file. This makes the process fail proof as the core concepts of Object Oriented Programming can be used to create an illusion of file using objects. The object readable files can be created using any languages as JAVA can produce SER files and Python can produce Pickle files. They are nothing but the objects containing information that are serialized and written into the above mentioned format of files such that again if needed to use the information, an object of the same kind which is used while writing it must be declared. They can not only act as tagger files but also, will

make the querying easy as they contain the gist, there is no need to search for the parent file but the tagged file contains enough information.

1.7 NLP in files

For building a querying mechanism, machine learning and Natural Language Processing(NLP) are the most predominant one to consider. As the way of searching must be a semantic one rather than syntactic. As the vocabulary is flooding with whole new words and new blended words are filling all over in the modern era, the capacity of people to remember the keywords present in the file becomes an improbable one. Thus the system must be able to map the keywords used for searching along with their meanings so that the result may be a desired one.

There are many ways to extract the important words that form a document using sophisticated machine learning algorithms like topic modelling, vectorization, page ranking etc. These are widely used algorithms in the field with page ranking, otherwise called TF-IDF (Term Frequency and Inverse Document Frequency) being the most successful one. The all known famous tech giants Google.inc published a paper on Page Ranking with its basics containing none other than TF-IDF is being used to fetch the links in Google search engine which works on accuracy of 99.672%.

1.8 Architecture design

The most notable point to perform while building these kinds of systems is choosing perfect architecture for it. Choosing the best architecture among the options of peer to peer(P2P), Client Server is the most vital one. Since there are more data to be stored to be used throughout the process like key, access list etc, it will be more advantageous to use Client Server architecture than P2P as there

is a need to keep an trusted authority to store the keys and encrypted files along with the tagged file, it will be meaningless to use P2P architecture.

CHAPTER 2

LITERATURE SURVEY

2.1 Overview

This chapter discusses the related work that is done in the field of security in file sharing, retrieval methods and many other complex mathematical models. It includes the different methods proposed so far.

2.2 Related Works

[1] The main work paper talks about is identity based encryption particularly on lightweight devices as the low powered devices need specific algorithms due to their low processing power it provides a client server type architecture. It manages the file in a mobile cloud computing environment with RSA (Rivest-Shamir-Adleman) being the primary encryption standard. The work also speaks about the Diffie Hellman algorithm that is primarily used for key sharing purposes. Identity Based Encryption is one of the key factors mentioned in the proposed system. The security analysis has been made comparing the proposed model with AES/CBC encryption standard with key size being 128-bit. The main architecture has a cloud controller which acts as the main purpose server and the users are divided into two kinds namely owner and consumer. A trusted authority called key generator is used to share keys between owners and consumers for encryption and decryption purposes. Bilinear pairing scheme is the main motive to provide degredency in the curve of productivity. The paper produced is just a theoretical model on securing file systems which have low computational power, the authors state as building a prototype being their primary future work.

[2] The following paper is a survey that has been published to potentially speak about the mobile cloud computing platform. It very highly provides the information regarding the developing cloud platform especially among the mobile devices and backs the information along with the statistics of papers cited. The authors mainly focused on the basics of what a mobile cloud computing is, its core concepts of architectural design and the pros and cons of adopting this kind of mobile environment. It also sheds the light on the areas that have opportunities available for research and also stands about the commerciality and marketing possibility as a whole. The main problem the authors agree upon is the security issue being marked along with the 24hr availability of cloud services as the energy and battery is what the whole world is searching to move to sustainable resources. The companies have started adopting a private cloud infrastructure and the information of people have reached a red alert. The paper concludes with stating the fact that more than two hundred million businesses will go on cloud platforms in and around 2025.

[3] The paper briefs about a novel method for password based key exchange among three parties. The main advantage that the work produces is the eradication of smart cards and server keys which is a conventional method that has been used so far for an authenticated key exchange process. Another novel scheme the authors introduced is the chaos algorithm, as well known chebyshev maps have been used since the 90's , with the disadvantage being predictable complexity and replay attacks, the new form of mathematical model that works on randomness in a non-linear robust system called Chaotic maps have been introduced. The authors made the research in a single server environment with an assumption of Chaos based decisional Diffie Helman(CDDH) and proposes that the whole scheme is a part of Random Oracle Model(ROM). The performance has been tabulated with the calculation of chebyshev polynomial

being approximately 32 milliseconds. The work is also compared with the research already made by Lai and Zhao, the overall performance has been compared with run time being the key factor and clearly this paper proved to work on less complexity by the margin of two milliseconds. The researchers clearly state to extend their work on a multi server environment with similar assumptions on CDDH.

[4] The research paper mainly schemed to showcase and shed a bright light in maintaining the privacy of critical data along with auditability. The authors brief about the integrated key management protocol that can be utilized thoroughly during emergency retrieval. The attribute based key generation is also a notable novel work produced in the paper stating the production of key with the attribute of the consumer such that more uniqueness can be achieved along with less predictability. The work is made on sensible health data. The key factor in the work is retrieval of data on emergency situations which is achieved through keyword based searching mechanism which not only preserves privacy but also hides the search pattern based on redundancy. The system model is assumed to be a public cloud where the data are being uploaded and an authority called Emergency Management Technician(EMT) is allotted and given access to all data pertinent to healthcare and only in emergency situations. The work assumes the EMT to be rational so that it becomes a trusted party. Searchable Symmetric Encryption (SSE) is adopted where the data are encrypted with complex data structures providing an easy retrieval mechanism. The paper also clearly states the future vision of authors being to devise a mechanism that completely provides integrity on data, which ensures whether any data has been illegally taken over or leaked and if leaked finding who is the disled authority.

[5] The research describes mobile health monitoring in cloud environments, with the encryption is done by Boneh Franklin's IBS (Identity based Encryption) and the model outsourcing the decryption techniques still preserving the privacy and maintaining the integrity of data. The work also relishes the newly proposed key private proxy re-encryption which is mainly adopted to reduce the computational cost of the system. The cloud servers are assumed to store the encrypted monitoring data and the presence of a semi-trusted authority (TA) is assumed. The TA is responsible for sharing keys to the mobile clients for the decryption process. New level of branching techniques is adopted and maintained that branches the input vectors that the TA needs to provide as a proxy key to both client and server for the encryption and decryption process.

[6] The work stands about the newfangled method for securing data, the cutting edge research reveals the introduction of a method called signcryption. The authors combined the mechanism of identity based encryption along with this newly proposed signcryption. The whole method is implemented using bilinear maps, which are as efficient as chaotic maps but lack the randomness, it stands ahead in providing less complexity with key size even being small. The main idea behind the research is to provide security along with non-repudiation. Usually the integrity of a message is maintained by signing the data after encryption but the signcryption gives a new method of providing encryption along with signing in a single step thus reducing the complexity.

As the steps are reduced with sign and encrypt in one step and decrypt and verify at other end, the scheme is compared with all the preexisting signcryption schemes which has been produced as research. The study revealed the work to be pretty much effective compared to all 6 existing papers, the difference in time taken for signing and encrypting the data takes much less time around two

milliseconds compared to all conventionally specified methods which lie between 5-6 ms. The decryption and verifying step was also way ahead of the competing papers as the time taken was only round 6 ms.

[7] The work described in the research paper enlightens a novel Identity based authentication scheme for data privacy preservation. Since there can be a conflict of interest between a cloud service provider and the consumer, a proper unbiased third party (TP) is taken into consideration for system variable transformation amount the service provider and consumer. The TP not only negotiates variables but also acts as a proper auditing tool to ensure the actual presence of data in the cloud for the service provider.

The attribute based encryption standard has a major disadvantage when the Key Distribution Center (KDC), is a single trusted authority. Though many fresh papers were published regarding the use of 0 to many KDC, this makes the load in the decryption process, especially while using mobile devices. Thus a better way to solve this crisis is using decentralized KDC also by outsourcing the decryption processes as mentioned in [4].

The whole mechanisms work under token based mechanisms, as the secret keys are transmitted along by the decentralized KDCs. The main use of the decentralisation is stopping the replay attacks which is a huge advantage compared to other schemes. Another positive side of this research includes the option of providing fault less multiple read and multiple write mechanisms which was very tedious to implement in recent research.

[8] This research is one of the most important findings and a cutting edge work on the security domain. The authors Boneh and Franklin produced a novel scheme of IBE scheme using bilinear pairing among the abelian groups that are found to be relatively of higher order. The proper high order group

mathematics has started to shine since the research of Weil pairing which is till now used in elliptic curve algorithms. The inability of predicting the keys in Elliptic Curve Cryptography(ECC) is due the randomness in the IBE using Diffie Helman proposed in this paper. The main two points specified along this is the Revocation of public keys, which allows the public key certificate to be valid for only the specified time frame and Delegation of decryption keys.

[9] The main idea behind the paper is to provide a secure data processing framework that addresses all the conventional problems in mobile cloud environments. The processing includes, storing, managing, retrieving the data. Since the mobile users have low computational power the architecture is built in such a way that each user device is mapped to Extended Semi Shadow Images(ESSI) and Service Nodes(SN). Thus every time when the user performs an operation, it is sent to ESSI and executed from ESSI such that it will screen the computational power of user nodes. ESSI along with SN are managed to form a Cloud Trusted Domain. The transmission of information among the user device and Trusted domain is performed through SSL and IPsec. The additional feature provided along is the management of multi tenancy in the same public cloud infrastructure.

The Authors promise on their future work on the investigation of tri-rooted ESSI solution along with auditing of data primarily with the misuse and data theft.

[10] The paper speaks about the heterogeneity in the cyber physical systems due to the difference in the physical embeddings along with the computational power and battery life. The main idea behind the research is to present a framework that is purely based on infrastructure which incorporates not only

structural but also semantic mappings to manage multi-model heterogeneous cyber physical cloud systems.

[11] The work presents the view of Term Frequency(TF) and Inverse Document Frequency(IDF) in the nature of text mining. These algorithms were devised to find the important words in corpus such that there will be no need for reading the whole document to know the gist, instead a few keywords can produce what the document mainly talks about.

It not only states the working of algorithms, it also states where to definitively use them such that it will be more effective and what are the environmental variables that affect the accuracy of the algorithm. It states why the algorithm fails in certain perspectives and devises a strategy to make use of the weakness.

[12] The open access paper talks about the classification of research paper, so that it will be advantageous for the readers to find the desired paper. The main idea used is Latent Dirichlet allocation (LDA), which tags the articles with topic vectors, then K-means clustering is used to cluster all the similar reference papers by their topic vectors, the semanticity is obtained through the TF-IDF scheme which then later used to scale the important words in a corpus. The future work author mentioned is expanding the application with multifarious applications like documents, tweets etc.

CHAPTER 3

PROPOSED WORK

3.1 Overview

This chapter briefs about the actual work done by categorizing the work into four divisions. It not only states the work done rather it establishes the importance of the work along with the reasoning of its existence.

3.2 Ideology

Security in computing is one of the most researched parts of this field around the globe. Hence in this proposed system we intend to produce a novel way of providing security to files in a server. As servers are meant to store sensible files like customer records which have more high end importance and must be leakage proof. Thus, providing a safer way to ensure its security results in the reliability of the people towards the platform. The thriving technologies push people to use more mobile devices which not only have low battery life but also less computation power. For these low powered devices to perform, the algorithms must be written considering the lightweight devices, as it is possible to write a layer of security algorithms for a super computer, but the actual tyranny comes while facing the lightweight devices. Hence the work to counter the disadvantages of the lightweight devices. The pre-written works were done on RSA algorithms, which is now expanded to modified ECC whose application and advantages will be briefed in the following chapters. Another most important aspect of storing data is retrieval, the work also incorporates an efficient querying mechanism, that allows the users to search for files by providing keywords. The work is further extended to an extent of images, where

before storing, the user must tag the image with keywords, which will be used to search in future.

3.3 USER REGISTRATION

3.3.1 Introduction

User registration is an important step to be done before providing any service. This is how the company that provides the service knows about the consumers and can troubleshoot the problems if necessary. It not only provides insights of the customers to the company but also acts as a facade to any malicious threat from an outdoor agent, like Denial-of-Service (DOS) attacks, bufferOverflow attacks etc.

3.3.2 Process

The users are divided into two categories namely the data owners and consumers. Owners are responsible for uploading the file with the accessibility restriction into a server and eventually an authorised consumer can search for the file and consume it, provided the consumer has given the access to the file by the owner. Each owner must create a group before uploading the file, and the consumer must specify the group he/she wants to be in.

Each owner can have one or more consumers, whereas a consumer can only register in one group at a time. All the services are maintained by session management techniques in order to handle distributed denial of service(DOS) attacks. In Order to utilise all the services, both the owners and consumers must register themselves in the server.

During the procedure of registration the users will be prompted to enter their details regarding themselves like EMAIL ID, PHONE NO, NAME,

ADDRESS,etc. Also they have to create login credentials including personal username and password for future use. For the registration of an owner, he/she must specify an UNIQUE ID from which the private key for encryption will be calculated and a GROUP NAME must be specified explicitly while prompted. Similarly while a consumer is registering, he/she will be prompted, but the a consumer should choose any one of the group they have to be a part of, which would be already specified by the owners

Thus at the end of the process, irrespective of an owner or consumer, both will have an USERNAME and PASSWORD. All the registration process is backed up in a database which would verify if any problem occurs.

3.4 MUTUAL AUTHENTICATION

3.4.1 Introduction

Authentication is a major checkpoint to perform while securing the systems. It validates the user as a non-malicious one and a trusted authority. The authentication must happen on both the ends in a client-server architecture, i.e, the client must be sure that he/she is in communication with the right server and the server must authenticate the client so that no malicious entity can intervene in the action of the server.

3.4.2 Process

As mentioned above, the importance of authentication in a client- server model, the work also incorporates a similar mechanism to ensure the security of the model.

As the owners and consumers register themselves to the service provider, the endproduct will be a username and product through which they can access

their profiles,i.e, owners can upload files and consumers can consume the file provided in the same group. The next step will be for the owner's to authenticate the consumer in their group.

At once a consumer selects the group to be a part of, an request prompt is sent to the owner of the group, stating that the particular consumer has requested to join the group. Authenticating the consumer to the group is completely in the hands of the owner. The owner can completely ignore the request of the consumer such that they will not be a part of the group and will not have access to the files.

When the owner authenticates the consumer to join the particular group, the process of key sharing occurs between them. A share of public key and private key will be provided to the consumer, such that he/she will be able to decrypt the files which are uploaded in the group.

Other than just sharing the keys among the clients, another cutting edge feature included is the control list. i.e, the owner can set the power a client can exhibit in a file. The owner can provide read, write or both read/write access to the client so that they have restricted ability on a file.

3.5 ENCRYPTION

3.5.1 Introduction

The above mentioned concepts like authentication and authorisation are just a facade for the inner security measures. The actual securing part is encrypting the file that is to be stored in the server. There are many types and levels of encryption and each has their own advantages and disadvantages. Before considering the algorithm to be used, the circumstances and external environment should be studied, here the game changing circumstance is the low

computational power of the devices which should be kept in mind before devising an algorithm.

3.5.2 Process

After the registration and the authentication of the owners and consumers, the next key part of the work is encryption. The file which may be a text or an image file (extensions include, .doc, .txt, .jpg, .png) can be fed to the server storage by an owner such that the file gets encrypted and stored in the storage in a distributed manner. The distributed manner includes, the files are split into one or many blocks and each block is encrypted and stored in a separate location, such that it will be impossible for an intruder to find the whole encrypted file to crack and decrypt it. The blocks are not stored in any particular order, as they are scrambled by a random number generator algorithm, and the actual order is updated in the database, such that during retrieval, it will be merged in the order it was scrambled earlier.

The next negotiable part is devising the algorithm for the process. Considering many factors including the battery life and low computational power there are many algorithms already proposed. RSA is one such algorithm which does the job especially in lightweight devices. RSA is a pretty old algorithm but still maintains its integrity in notable situations. Still the SSL/TLS (Transport Layer Security) for a browser still uses RSA in the backend..

RSA is a declared ROM (Random Oracle Model) algorithm which uses trapdoor functions to encrypt and decrypt data. The actual concept used is multiplication and discrete logarithms. Two large prime numbers are multiplied and the resulting public key is used to encrypt the data. The encrypted data is

sent to the receiver, where they use a private key to decrypt and get back the actual message. If any malicious agent intercepts the data along with the public key, in order to decrypt the agent should factorize the product into two bigger primes to find a private key which is mostly considered as an improbable task.

But due to the enhancement of technology and introduction of quantum computers, many parallel and distributed computing researchers have found new algorithms like gradient sieve etc, to crack the factorization problem. Though the process can be made hard by choosing larger prime numbers, it will break at some extent.

Then to overcome the disadvantage of RSA, elliptic curves are studied and a new idea has been developed to use the curves to encrypt data. The work thus incorporates the Elliptic Curve Cryptography(ECC) as the encryption algorithm.

The ECC is similar to RSA in encryption and decryption with public and private keys, but the ECC uses coordinate points in an XY Plane as keys. Another mathematical lemma it incorporates is the reflection of points theory, a public key is chosen as a point and reflected across the axes for a certain number of times to get a private key. As it is impossible to track the trajectory of reflections, it is impossible to find the private key by any intruder.

3.5.3 Countering the disadvantages

The only attacks that can be performed in the system are Brute Force attack and Power analysis attack.

In power analysis attack the middle man tries to find the time taken by the algorithm to decrypt a file, and from that analysis guesses the key size in bits. After knowing Key Size, brute force attack will be performed to guess the key

used. To overcome this, one way is to speed up the calculation or make the time constant for all operations.

To make all the operations take a constant time to complete $O(n)$, where n is an integer, an algorithm called Fast Montgomery Ladder is written. And to increase the efficiency in operations, a mathematical model called Comb Multiplication is incorporated. The comb multiplication reduces the redundant steps of a compiler and replaces them with bitwise operations such that the execution time will be faster.

3.5.4 Reason for ECC

By research it is found that the security provided by the 190bit key of the ECC is equal to the 2048bit key used in RSA.

3.6 FILE RETRIEVAL

3.6.1 Introduction

Files are stored in a safer place so that it can be accessed without any discrepancy. Due to the increase in needs of securing files on the internet, there is a need for a better algorithm to retrieve the files when needed. The algorithm should be devised by keeping an eye on both accuracy and time complexity in hand.

3.6.2 Process

Once the owner uploads a file for a group, it gets encrypted and stored in the filesystem as blocks in a scrambled manner as specified above. There will arise a situation when the owner uploads many files in a group and the consumer will have to search and retrieve the file he wants manually.

To overcome this disadvantage, an efficient multi keyword search algorithm is incorporated for every client. The facility can be used to search the files present in the file system by specifying keywords related to the document to be found.

3.6.3 File Preprocessing

To make the querying of the files possible, a certain amount of preprocessing is done to the files before they are encrypted and uploaded. The files are read once and a tagger file(Configuration File) is generated for every uploaded file and the tagger files are stored with the encrypted file in the filesystem. The Configuration files contain the keywords extracted from the actual file that form a gist of the parent file. The tagger file also known as the configuration file, contains enough keywords mapped along with their semantic meaning and the root words to act as a proxy of the parent file.

Since the tagged file is also present along with the parent file in the file system, the security is provided by making it an object readable file. i.e, it cannot be directly visible directly rather it will only be accessible through application codes. In brief the text mining algorithm is made to run over the corpus being uploaded and the keywords mined from the corpus are stored in an object. The object is later serialized and written as a .SER file.

Hence whenever a consumer searches for a file using keyword, the SER files are searched and if the keywords match, then the parent file is returned in a decrypted manner.

3.6.4 Text Mining

Another notable point the work established is the incorporation of text mining algorithms. Many conventional algorithms were tried to extract

keywords from a corpus, like Topic modeling using Gensim, where the corpus is tagged with topics that resembles the document, another is TF-IDF (Term Frequency and Inverse Document Frequency) which proved to be a much effective one than the latter as they take each and every words into consideration before ranking them upon their importance.

3.6.5 Image Files

For creating configuration file images which are being uploaded, running sophisticated object identification algorithm will provide a better result in tagging the image, but taking into consideration the low computational power of lightweight devices and complexity of algorithms in JAVA, the application will prompt the user to tag the image with specific keywords manually, later before serializing them into SER files, the keywords entered are mapped along with their root words and other words in vocabulary using WORDNET interface that resemble the entered keyword.

Finally when a client is searching for a file with a keyword, the results will be a mixture of both image files and corpus that resembles the keyword used.

3.7 ARCHITECTURE DIAGRAM

3.7.1 Overview

The diagram below(fig 3.1) clearly states the architecture of proposed work which is clearly explained above.

3.7.2 Diagram

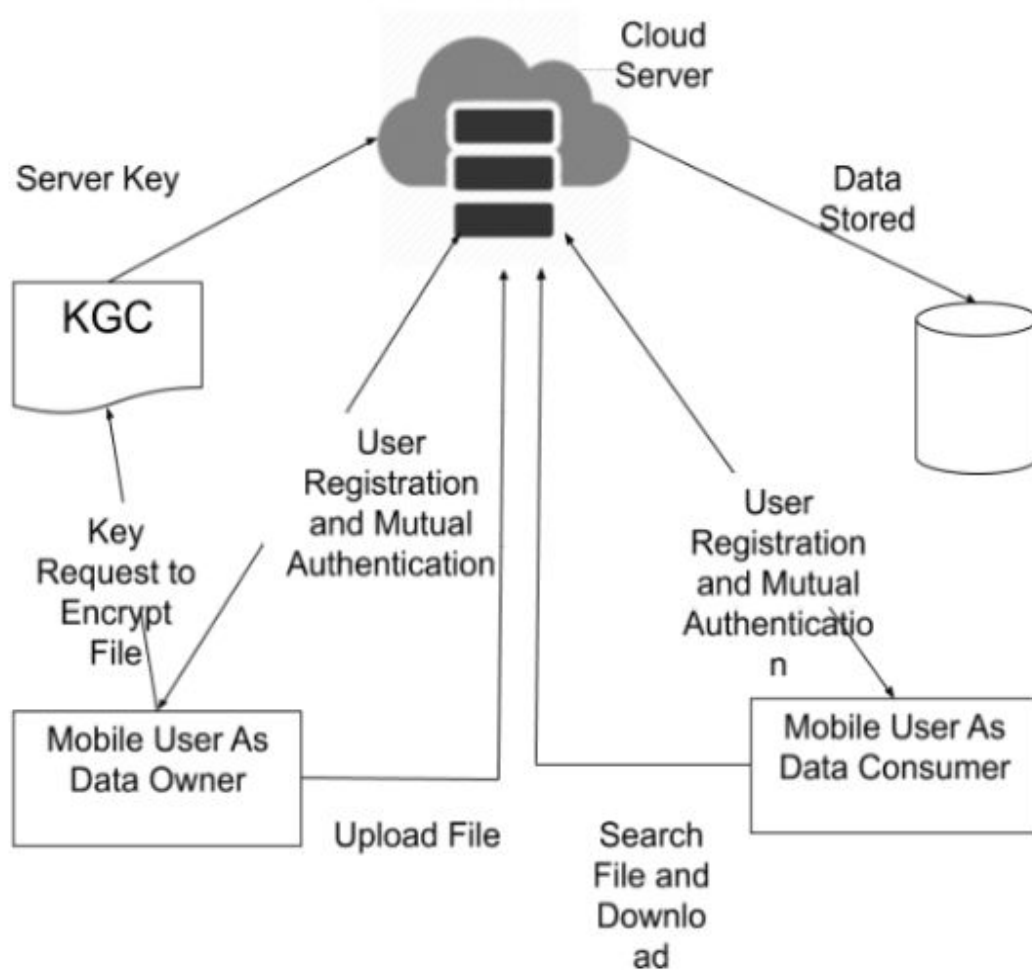


Fig 3.1 Architecture diagram for Secure File Sharing in CPC

3.7.3 WorkFlow Diagram

The diagram(fig 3.2) is to explain the workflow for the work proposed above

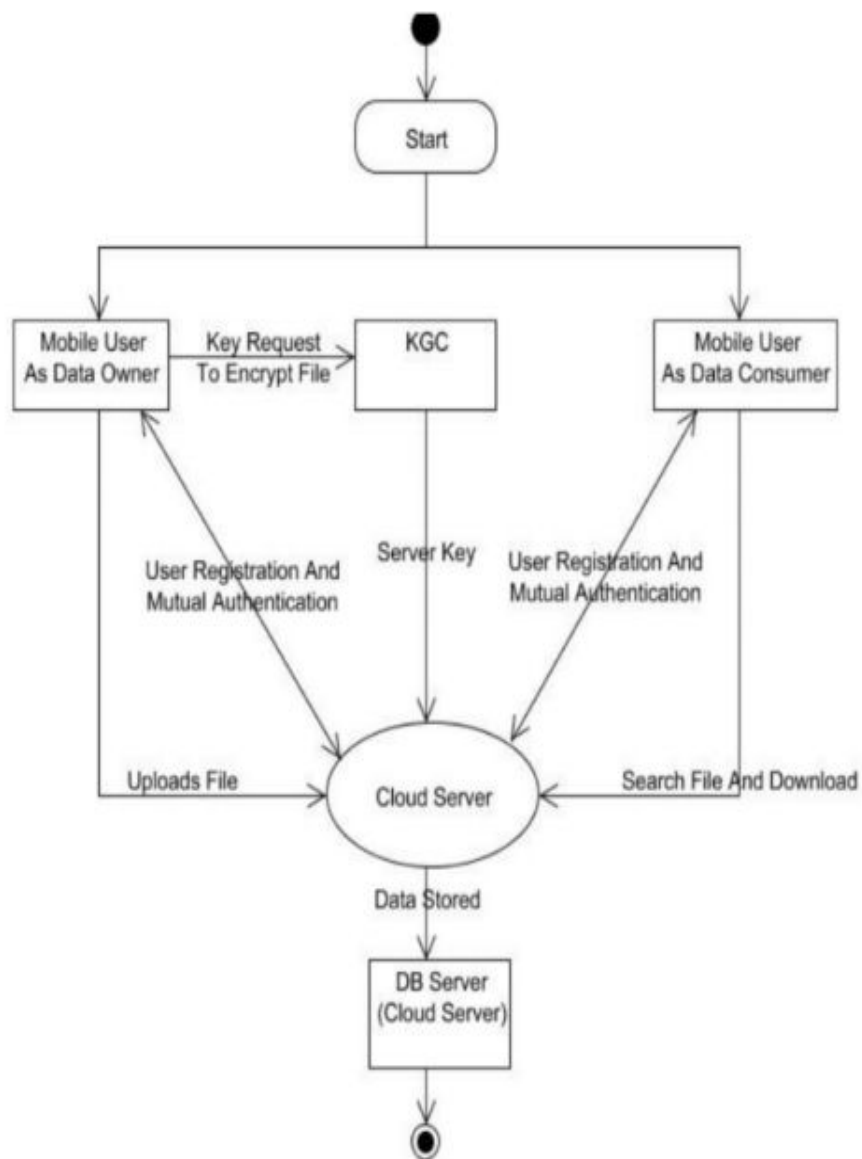


Fig 3.2 - workflow diagram

CHAPTER 4

IMPLEMENTATION DETAILS

4.1 Overview

This chapter briefs the algorithms used and their purpose in the work with explanation and examples. It states the reasons for assumptions made and the reason for choosing to do each step.

4.2 Mutual Authentication

As stated the importance of the authentication in securing files, the key sharing mechanism is briefed in the following subdivision. The chosen algorithm to share the key among servers and clients is Diffie-Helman. The algorithm comes under the paradigm of random oracle models, hence a reliable one. The algorithm works on trapdoor functions, where the complexity of doing entry operation is very low, whereas reversing the same process takes high level complexity, and even improbable for larger numbers with bit size more than 4096.

4.2.1 Diffie-Helman

The main purpose of using Diffie-Helman is to share the keys between two parties.

Assumptions : The line is insecure, which can be intercepted

Algorithm :

AIM : To share the keys among users A and B
Environment : A and B are connected through unsecured line of communication
Reason : Practicality is maintained considering a malicious interceptor

can obtain the information

Step 1 : P be a very large prime (not decomposable and more than 2048bits)

Step 2 : G is assumed as primitive root, condition : $G < P$

Step 3 : 'a' is assumed as private key of user A, condition : $a < P$

PUBLIC KEY OF USER B CAN BE CALCULATED BY 'a'

Step 4 : 'X'(Public key of B) = $G^a \bmod P$

Step 5 : 'b' is assumed as private key of user B, condition : $b < P$

Step 6 : 'Y'(Public key of A) = $G^b \bmod P$

PUBLIC KEYS ARE SHARED THROUGH INSECURE MEDIUM

Step 7 : User A will have {a,X} ; User B will have {b,Y}

KEYS CAN BE CALCULATED BY

Step 8 : User A calculates K1, $K1 = X^a \bmod P$

Step 9 : User B calculates K2, $K2 = Y^b \bmod P$

Step 10: The keys obtained will be the same, i.e, $K = K1 = K2$

Step 11: Now users can use the symmetric key K obtained for encryption

REASON FOR SAFETY : The line is considered as an insecure, even an interceptor knows G, P, X, Y , the malicious agent will not be able to obtain the keys without knowing the private key of the users. Hence this method is prone to only brute force attack, where the hacker must try every number in order to obtain the keys.

4.2.2 Conclusion

Thus, the essentiality of key sharing is explained along with the reason of safety, although there are many key sharing algorithms, Diffie-Helman proved to be made for lightweight devices due to its less complexity in execution.

4.3 ENCRYPTION

4.3.1 Introduction

The core part of the framework is encryption and decryption. The keys are generated and shared among the clients which will be used for encryption by the server and decrypted by clients. There are many algorithms that can be used in this part, considering the working environment of this product, Elliptic curve cryptography is used.

4.3.2 Why ECC? Why not RSA

Both the considered ECC and RSA are public key cryptography methods, the key sharing part is fortified by Diffie-Helman, the former proves better than latter in their key creation methods.

The keys generated by RSA are two prime numbers multiplied with each other. Though the factorization back to the two numbers is difficult, it is not impossible, due to the rising demand in technologies. It can be made possible by supercomputers and quantum computers using methods like parallel computing.

The keys generated by ECC are rather points in the coordinate plane obtained by point reflection theory. The security that is provided by a 190-bit ECC key, is equal to security provided by a 2048-bit RSA key.

4.3.3 Elliptic Curve Cryptography

4.3.3.1 Assumptions

1. Elliptic curves are bi-linear coordinate curves whose points are defined by equation :
 - a. $Y^3 = X^2 + aX + b$
 - b. 'a' and 'b' are deciding parameters of curve
 - c. condition : $4a^3 + 27b^2 \neq 0$ (i.e, curves must be singular)

2. The elliptic curves are symmetric about X - axis
3. An ideal point called point of infinity is considered and denoted by symbol 0 (ZERO)
4. Let G be a group, for all $P \in G$ and let $\{P_1, P_2, \dots \mid P_i = (x, y) \text{ are points on a coordinate plane} \}$
 - a. Identity element 'i' is point of infinity 0
 - b. Inverse of a point is its reflection about both the axis
5. Point_Addition ($G, +$) is a rule defined over group G :
 - a. For any three non-zero, aligned element P, Q, R of a group G , the sum is given by : $P + Q = -R$
 - b. If $P = 0$ or $Q = 0$, certainly no line can be drawn to find the reflection over an axis (as lines are parallel). As 0 is already defined as an identity element, the rule can be redefined as $P + 0 = P$ and $0 + Q = Q$ for any Element P and Q of group G .
 - c. If $P = -Q$, the line going through points is vertical (Perpendicular), hence does not intersect at any third point. But if P is inverse of Q , then by inverse definition, $P + Q = 0$ for any P and Q of group G .
 - d. If $P = Q$, infinitely many points can be drawn through the points, by addition rule, $P + P = -R$.
6. Unlike point addition many rules can be defined over group G like :
 - a. ALGEBRAIC ADDITION
 - b. SCALAR MULTIPLICATION
 - c. INVERSE

The below figure (Fig 4.1) depicts the logical point distribution in an elliptic curve.

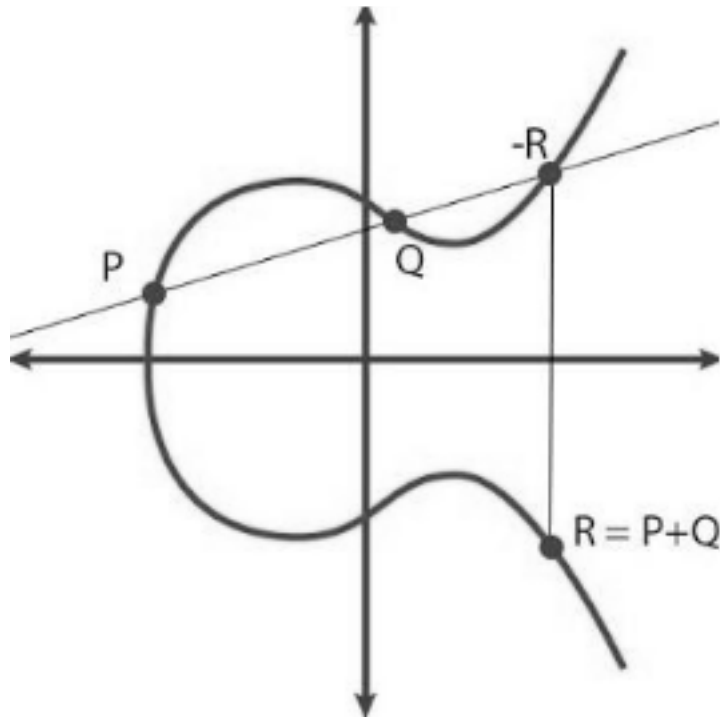


Fig 4.1 - Elliptic curve Point representation

4.3.3.2 ECC - Point Generation

1. Steps defining group operation

AIM : Define operation in group G

1. Let G be a group, for all $c \in G$ and let $\{C_1, C_2, \dots \mid C_i = (x, y) \text{ are points on a coordinate plane}\}$
2. Let **Point_Addition** : operation $(G, +)$ be defined on a Group G :
 - a. if $P, Q, R = (x, y) \in G$ and lie in same line
 - i. **Point_Addition** $(P, Q) = -R$
 - b. if P or Q does not exist in boundary and i is identity element :
 - i. Then resulting point considered is i
 - c. if $P = Q = (x, y)$ are same and due to (a) :
 - i. **Point_Addition** $(P, P) = -R$ (inverse of resulting point)

also,

- ii. $\text{Point_Addition}(P, R) = -P$ (inverse of generating point)
- d. if P is inverse of Q, i.e, $P(x, y) = Q(-x, -y)$
 - i. Q is called reflection of P with respect to both axis
 - ii. $\text{Point_Addition}(P, Q)$ results in a point at extreme(infinity)

3. Let INVERSE : operation $(G, -)$ be defined on a Group G :

- a. if R belongs to G
 - i. $-R$ is reflection of point R

4. Let Scalar_MUL : operation (G, \cdot) be defined on a Group G :

- a. if $P = (x, y) \in G$, then $n \cdot P$ where $n \in Z$ is defined as
 - i. $\sum_{i=1}^n P_i$ (Where summation done on Point_Addition algorithm defined over group G)

2. Steps defining point generation

AIM : To finds points on a curve

- 1. Let $P = (X, Y) \in G$, where $G = \{s_i = (x_i, y_i) \mid i \in Z, s \text{ is points on elliptic curve} \}$
- 2. To find another point on the curve (Q) as G is defined as abelian group
 - a. $P + P(Q)$ is performed
 - b. For each new point Q generated perform :
 - i. $P + Q$ to get new point(Q_{new})
 - ii. until Q_{new} not belongs to G

4.3.3.3 ECC - Key Generation

AIM : To generate Private and Public keys for encryption and decryption purpose

1. Order $|G| = N_G$ is efficiently calculated by Schoofs algorithm
2. Let $G_s \leq G$ and $|G_s| = N_g$
3. Cofactor $c = N_G / N_g$
4. Let $P=(x,y) \in G_s$ and since $G_s \subset G$:
 - a. $*$ (SCALAR_MUL) is defined on G_s
 - b. Hence , $BP = c * P$
5. Private key $PR_KEY = \text{RAND}(1, n-1)$
6. Public key $PU_KEY = PR_KEY \cdot BP$ (BP is a base point)

4.3.3.4 ECC - Encryption and Decryption

AIM : To perform encryption and decryption using keys generated

1. Assume K be a message(Confidential) on the curve as any Point P.
2. Key generator is the random number(d) that is formed from the range $[1, n-1]$ for encryption.
3. The new message(I,J) formed has a pair of points combined to form cipher text.
4. $I = d * G$
5. $J = P + (d * G)$
6. Multiply I and Private key
7. Subtract the multiplication value from other point(J)
8. $P = J - (dJ * I)$

9. Normal text H is found

4.3.3.5 ECC - Montgomery Ladder

Introduction

The proposed method of ECC is vulnerable to attacks namely :

1. Power Analysis Attack
2. Brute Force Attack

The power analysis attack is a way of predicting the time taken by an algorithm to encrypt and decrypt a file and through which predicting the key size to perform brute force attack. It can be countered by making run time constant for any key size through Montgomery Algorithm

Montgomery Ladder

Step 1: Let K be the factor such that $\gcd(N,K)$ is 1 and $N > K$.

Step 2: Then calculate the following values

- a. $N^{-1}(\text{Inverse of } N) \pmod K$
- b. $x! = x * K \pmod N$
- c. $y! = y * K \pmod N$
- d. $z! = (x! * y!)^{K-1} \pmod N$
- e. $z = z^{K-1} \pmod N$

Step 3: $z = x * y \pmod N$

- f. $z! * K^{-1} = (K-1)(x! * y!)(K-1) = x * y \pmod N$

4.3.3.6 ECC - Comb multiplication:

Introduction

Point multiplication plays a critical role in ECC. Many efficient and complex mathematical models have been proposed to ease the procedure of multiplying points in a coordinate plane such that the complexity remains low. One of them is the comb method which is much more efficient than other methods if precomputation points are calculated in advance or elsewhere.

A scalar multiplication method which uses pre computations for a fixed point P , meaning that they precalculate and store values that solely depend on P to speed up a scalar multiplication performed later on.

It composes of many variants like

1. Fixed-base Comb multiplication
2. Sakuhai method for double and add
3. Lopez and Dahub method of differential Montgomery ladder

Fixed Base Comb Multiplication:

1. Proposed by Lim and Lee for scalar multiplication on points.
2. The scalar ' k ' to be multiplied is considered in binary form of length ' l '
3. The representation is considered as a matrix of ' d ' columns and ' h ' rows
4. The binary form is split into ' r ' blocks with each of length $\lceil l/d \rceil$
5. The size is adjusted to mentioned block size by padding with zeros in left
6. After splitting into blocks, each blocks are represented in rows of a matrix
7. To gain a speedup later on, it is necessary to precompute all possible bit permutations for a bitstring $s = (b_{r-1}, \dots, b_1, b_0)$ of length r .

Algorithm: Lookup table creation - Fixed-base comb method for point multiplication

Input : Generator point A , maxlen ,blocksize

Output: Lookup table LT of coordinate type

Step 1: Initialize LT as null

Step 2: Nblocks = maxlen / blocksize

Step 3: tableSize = $1 \ll n\text{Blocks}$

Step 4: Q=A

Step 5: for i=0 to nBlocks:

a. for j=0 to tableSize

i. If (j & ($1 \ll i$))

1. $LT[j] = \text{add}(LT[j], Q)$

b. for k=0 to tableSize

i. $Q = \text{add}(Q, Q)$

Step 6: return LT

Algorithm : Fixed-base comb method for point multiplication

Input : Window width w, $n = \lceil l/w \rceil$, $k = (k_{l-1}, \dots, k_1, k_0)$, $P \in E(F_q)$

Output: $k \cdot P$

1. Precomputation: compute $[S_{w-1}, \dots, S_2, S_1, S_0] P$ for a bitstring S of length w

2. for i = n - 1 to 0 :

```

a.  $Q = 2 * Q$ 

b.  $Q = Q + \sum_{j=w-1}^0 K_{j,i} * P.$ 

c. end for

3. return (Q)

```

4.4 Configuration File Creation

The core part in retrieval of an encrypted file is generating a configuration file. A configuration file is a key file that gets searched first before downloading the document. Whenever a user searches for a document with a keyword, the entered keyword is matched with the configuration files, and the files made up such keywords are returned to the user and the user can download it provided he/she has access to it. Thus, a considerable amount of preprocessing is done to file before it gets encrypted and uploaded. The whole file is read and a process called text mining is made to extract the highly important words that make up the document.

Those words are then encapsulated in an object and the object is serialized into a file of .ser format. This not only provides faster access while searching using a JAVA object, but also provides non-readability to a certain extent. Thus even though the files get exposed it cannot be read by a normally without using the proper object structure that is used to create the serialized form. The creation of configuration is made through a traditional algorithm called Term Frequency and Inverse document frequency. The most infamous method of Identifying the words that make up a document is TF_IDF. The main concept is that the weight of each neuron increases as the word frequency in a document increases.

Before devising the algorithm to feed in a corpus the following preprocessing are done.

1. Stop words removal - As they provide no meaning in defining a document
2. Unprecedented root words removal - As the deprecated words may make the mode weak in classifying.

ALGORITHM : TF IDF

Step 1: TF_IDF (T,D) is given by :

- $tf(T,D) * idf(W,C)$ {T:term in document , D: Document, C:Corpus, W: words in corpus}

Step 2: TF(T,D) i.e, term frequency given by:

- $f_{T,D}$ {f: frequency(no.of occurance of word T in document D)}

Step 3: IDF(W,C) i.e, Inverse Document Frequency given by:

- $\log \frac{S}{|\{c \in C : w \in c\}|}$
- S denoted the total number of corpus considered
- Denominator explains the number of documents that are considered in a data set that contains the word w.

CHAPTER 5

EXPERIMENTAL RESULT

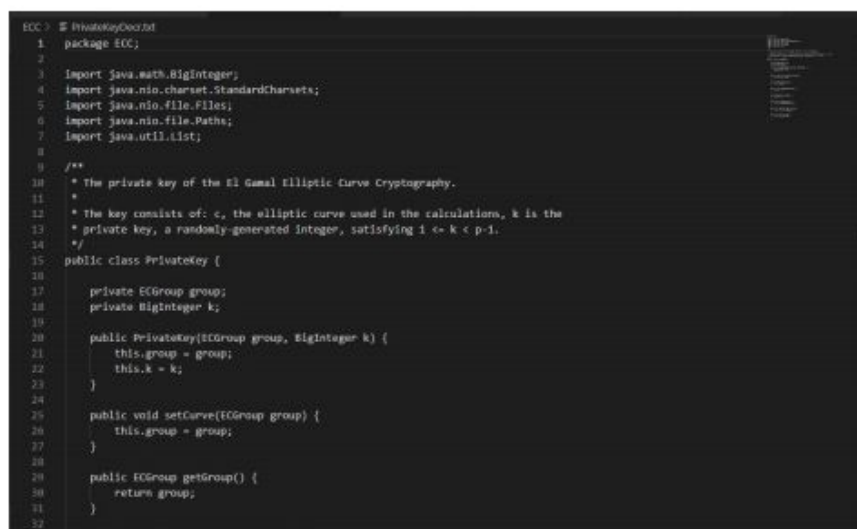
5.1 Overview

This chapter shows the results obtained at various stages of the project, with the images. The results are displayed as screenshotted images and pasted in JPEG formats.

5.2 Encryption and Decryption

INPUT :

The image below(fig 5.1) is the a java code, which itself is fed as an input to the encryption algorithm :



```
1 package ecc;
2
3 import java.math.BigInteger;
4 import java.nio.charset.StandardCharsets;
5 import java.nio.file.Files;
6 import java.nio.file.Paths;
7 import java.util.List;
8
9 /**
10  * The private key of the El Gamal Elliptic Curve Cryptography.
11  *
12  * The key consists of: c, the elliptic curve used in the calculations, k is the
13  * private key, a randomly-generated integer, satisfying 1 <= k < p-1.
14  */
15 public class PrivateKey {
16
17     private ECGroup group;
18     private BigInteger k;
19
20     public PrivateKey(ECGroup group, BigInteger k) {
21         this.group = group;
22         this.k = k;
23     }
24
25     public void setCurve(ECGroup group) {
26         this.group = group;
27     }
28
29     public ECGroup getGroup() {
30         return group;
31     }
32 }
```

Fig 5.1 - Input for ECC encryption

The figure below(Fig 5.2) is the output of encryption algorithm

Output:

[illegible]

Fig 5.2 - File encrypted by ECC

5.3 Configuration File Generation

INPUT :

The image(fig 5.3) is a text file chosen for configuration file generation

File Edit Format View Help

Donald John Trump (born June 14, 1946) is the 45th and current president of the United States. Before entering politics, he was a businessman and television personality. Trump was born and raised in the New York City borough of Queens, and received a B.S. degree in economics from the Wharton School at the University of Pennsylvania. He took charge of his family's real-estate business in 1971, renamed it The Trump Organization, and expanded its operations from Queens and Brooklyn to the Caribbean. The company built or renovated skyscrapers, hotels, casinos, and golf courses. In the 1980s and 1989, following setbacks in several highly leveraged real estate ventures, Trump diversified into various side ventures, mostly by licensing his name. He co-authored several books, including *The Art of the Deal*. He owned the Miss Universe and Miss USA beauty pageants from 1996 to 2015, and produced and hosted *The Apprentice*, a reality television show, from 2003 to 2015. Forbes estimates his net worth to be \$3.1 billion. Trump entered the 2016 presidential race as a Republican and defeated 36 other candidates in the primaries. His political positions have been described as populist, protectionist, and nationalist. He was elected in a surprise victory over Democratic nominee Hillary Clinton, although he lost the popular vote. He became the oldest first-term U.S. president, and the first one without prior military or government service. His election and policies have sparked numerous protests. Trump has made many false or misleading statements during his campaign and presidency. The statements have been documented by fact-checkers, and the media have widely described the phenomenon as unprecedented in American politics. Many of his comments and actions have also been characterized as racially charged or racist. During his presidency, Trump ordered a travel ban on citizens from several Muslim-majority countries, citing security concerns; after legal challenges, the Supreme Court upheld the policy's third revision. He enacted a tax-cut package for individuals and businesses, which also rescinded the individual health insurance mandate and allowed oil drilling in the Arctic Refuge. He appointed Neil Gorsuch and Brett Kavanaugh to the Supreme Court. In foreign policy, Trump has pursued an America First agenda, withdrawing the U.S. from the Trans-Pacific Partnership trade negotiations, the Paris Agreement on climate change, and the Iran nuclear deal. He recognized Jerusalem as the capital of Israel, imposed import tariffs triggering a trade war with China, and started negotiations with North Korea towards total denuclearization. Trump and his advisers of his 2016 campaign were suspected of being complicit in Russian election interference that was illegal, but a special counsel investigation did not find sufficient evidence to establish conspiracy or coordination with Russia. Trump was also personally investigated for obstruction of justice, and was neither indicted nor exonerated. Separately, in September 2019, the House of Representatives initiated an impeachment inquiry alleging abuse of office for political gain. In July 2019, Trump had asked Ukraine to investigate Hunter Biden, whose father Joe Biden is a potential rival presidential candidate for 2020, and a whistleblower alleged that the White House had tried to cover up Trump's request.

Fig 5.3 - Input text file for Conf file generation

The image(fig 5.4) is the output configuration file generated for the text file

Fig 5.4 - Generated Configuration File

CHAPTER 6

CONCLUSION AND FUTURE WORK

The proposed work not only guarantees the effectiveness in securing a file especially for lightweight devices but also provides an active querying mechanism to retrieve the file using keyword search feature. The project currently works perfectly on files containing texts, as a future work this can be extended by incorporating image recognition algorithms to separate and classify frames in a video files files so that they can be also encrypted and uploaded in the server.

REFERENCES

1. A. Karati, R. Amin, S. K. H. Islam and K. R. Choo, "Provably secure and lightweight identity-based authenticated data sharing protocol for cyber-physical cloud environment," in IEEE Transactions on Cloud Computing. Doi: 10.1109/TCC.2018.2834405
2. Dinh Thai, Hoang & Lee, Chonho & Niyato, Dusit & Wang, Ping. (2013). "A survey of mobile cloud computing: Architecture, applications, and approaches". Wireless Communications and Mobile Computing. 13. 10.1002/wcm.1203
3. Farash, Mohammad & Ahmadian Attari, Mahmoud. (2014)." An efficient and provably secure three-party password-based authenticated key exchange protocol based on Chebyshev chaotic maps." Nonlinear Dynamics. 77. 10.1007/s11071-014-1304-6.
4. BAMANE, Amruta et al." CLOUD ASSISTED MOBILE ACCESS OF HEALTH DATA WITH PRIVACY AND AUDITABILITY". International Education and Research Journal, [S.l.], v. 1, n. 5 , p. 18-20, dec. 2015. ISSN 2454-9916.
5. H. Lin, J. Shao, C. Zhang and Y. Fang, "CAM: Cloud-Assisted Privacy Preserving Mobile Health Monitoring," in IEEE Transactions on Information Forensics and Security, vol. 8, no. 6, pp. 985-997, June 2013
6. Barreto, Paulo & Libert, Benoît & McCullagh, Noel & Quisquater, Jean-Jacques. (2005). "Efficient and Provably-Secure Identity-Based Signatures and Signcryption from Bilinear Maps". Advances in Cryptology. 3788. 515-532. 10.1007/1159344728.
7. R. Vishnu Sekhar, N. Nandini, D. Bhanumathy, M. Hemalatha. "Identity Based Authentication for Data Stored in Cloud". International Journal of Advanced Research in Computer Science and Software Engineering(2015).
8. Boneh D., Franklin M. (2001) Identity-Based Encryption from the

Weil Pairing. In: Kilian J. (eds) Advances in Cryptology — CRYPTO 2001. CRYPTO 2001. Lecture Notes in Computer Science, vol 2139. Springer, Berlin, Heidelberg

9. Huang, Dijiang & Zhou, Zhibin & Xu, Le & Xing, Tianyi & Zhong, Yunji. (2011). Secure data processing framework for mobile cloud computing. 614 - 618. 10.1109/INFCOMW.2011.5928886.
10. Rajhans, Akshay & Bhave, Ajinkya & Ruchkin, Ivan & H. Krogh, Bruce & Garland, David & Platzer, André & Schmerl, Bradley. (2014). Supporting Heterogeneity in Cyber-Physical Systems Architectures. IEEE Transactions on Automatic Control. 59. 3178-3193.
11. Qaiser, Shahzad & Ali, Ramsha. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications. 181. 10.5120/ijca2018917395.
12. Kim, S., Gil, J. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019). 10.1186/s13673-019-0192-7.