Design of a Secure Privacy Preserving Cloud Based Sharing Platform for Electronic Health Data

by

Munachiso Ilokah

A thesis submitted in partial fulfillment
of the requirements for the degree of

Masters of Applied Science

in

Faculty of Engineering and Applied Science

Electrical and Computer Engineering

University of Ontario Institute of Technology

Supervisor: Dr. Mikael Eklund

July 2019

Abstract

The abstract will be rewritten at the end and should provide a complete narrative of the thesis i.e like a summary. The current abstract is completely focused on the ABE element while the focus of the thesis is on a framework for the secure sharing of health data through third party storage services such as cloud computing.

The need to protect user data from unauthorized access and malicious use by authorized users, especially in the case of a system that includes the use of a third party storage service, which limits the amount of control that the data owner has, continues to present itself. The use of encryption for secure data storage has continued to evolve in order to meet the need for flexible and fine grained access control which led to the development of Attribute Based Encryption (ABE) based on the concept of Identity Based Encryption (IBE). ABE gives more control to the data owner and has continued to evolve to enable users make use of the technological advancements available to them.

The use of ABE to ensure the security and privacy of health data has been explored with multiple solutions proposed. This thesis aims to develop a platform that applies an improved ABE scheme which allows for the secure outsourcing the more computationally intensive processes for data decryption to the cloud servers, reducing the

amount of time needed for decryption to occur at the user end and also reducing the amount of computational power needed by users who require access to the data stored in the cloud in its encrypted form.

Acknowledgements

- Acknowledge supervisor

- External academic support or guides

- External professional support (slc, ogs staff)

- Colleagues (lab mates, lokendra,etc)

- Personal Support (Liz, Tim, Almey, Tessa, Catherine, Margot, Aida, etc)

- Family (Mum, Dad, Brothers, etc)

Contents

List of Figures

List of Tables

List of Abbreviations

stopped at background and lit review chapter. also look into what abbreviations should be contained in the list i.e the important and critical one vs. all of the acronyms

IT	Information Technology
ABE	Attribute Based Encryption
CP-ABE	Ciphertext Policy Attribute Based Encryption
KP-ABE	Key Policy Attribute Based Encryption
MACP-ABE	Multi Authority Ciphertext Policy Attribute Based Encryption
PII	Personal Identifiable Information
NIST	National Institute of Standards and Technology
EHR	Electronic Health Record
EMR	Electronic Medical Record
PHR	Personal Health Record
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
IBE	Identity Based Encryption
CSP	Cloud Service Provider

CA Certificate Authority

Chapter 1

Introduction

1.1 Motivation

Secure storage of user electronic health data by institutions such as hospitals are important in order to enable them make this data available to other parties while ensuring that they prevent unauthorized access to user data and protect the privacy of their patients. These institutions typically store this data on physical hardware in a secure location on their premises and so are able to prevent unauthorized access through the use of adequate network security infrastructures and also prevent authorized users from gaining access through the use of proper access control mechanisms which ensures accountability. The deployment of more efficient and modern information technology infrastructures such as cloud computing extends the trust boundary of the Information Technology (IT) infrastructure of institutions beyond their facilities while giving them multiple advantages such as a potentially unlimited amount of storage and computational capabilities. This new infrastructure exposes data to

both internal and external threats as the institutions no longer have physical control of the infrastructure on which their data is stored as that is now in the control of the Cloud Service Providers (CSP).

One way of ensuring that user data is protected from both the CSPs and from unauthorized users is to encrypt the data before it is stored in the cloud using available encryption schemes. This allows only authorized users with access to the required information which is usually a key to gain access to data in its original form. This system becomes a challenge when there is a need to provide fine-grained access to data to third parties in order to ensure that privacy of users are not violated in relation to privacy laws such as PHIPA [1] which governs the collection, use and disclosure of personal information in the health sector in the province of Ontario, Canada. Sharing of data with third parties such as universities and research institutes becomes a challenge as the health institutions need to ensure that privacy of their patients are not violated while they provide these parties with the data necessary for conducting research which could lead to advancements in the health industry and improve the service delivery for patients while saving more lives.

The need to meet the privacy and security requirements have led to the creation of systems for the sharing of electronic health data based on a system of encryption called Attribute Based Encryption (ABE) [2] [3] [4] [5] [6] [7] [8] [9]. ABE, developed by Sahai and Waters [10], provides the opportunity for fine-grained access control to data as either the user secret keys or the corresponding ciphertexts contain descriptive attributes with the other containing an access structure that specifies what attributes need to be present for decryption. This means that, if there is no match between the attributes of a user's key and the ciphertext, decryption is prevented. This enables the health institutions to be able to grant third parties access to only the specific

data that they require to carry out their research while ensuring that Personal Identifiable Information (PII) of patients are not exposed in line with privacy laws and requirements.

1.2 Thesis Contributions

To be revised i.e rewritten.

In this thesis we have developed a platform for the sharing of health data among multiple parties using untrusted third party storage services. The platform employs the use of symmetric encryption for the security of the actual data sets while a variant of the Multi-Authority Ciphertext Policy Attribute Encryption (MACP-ABE) is applied to encrypt the secret keys that users need to gain access to data in its plain form. This is because of the amount of computation required for encryption and decryption algorithms of the different available Public Key encryption schemes, a category that includes the MACP-ABE scheme. The security of this protocol had been proven adequately in the original work [cite original paper] - the original dac mac scheme being modified and as a result, we have limited our proofs to the basic requirements such as correctness of the encryption scheme.

The major contributions of this thesis are -

- Decryption Time (at User End) - a reduction in the amount of time on the end of the user for the decryption of data
- Reduced Computational Requirements - the computationally intensive operations required for decryption are offloaded to the cloud servers, thereby reducing

the computational requirements for end user devices for data decryption

1.3 Thesis Outline

To be revised i.e rewritten.

The rest of this thesis is organized as follows: Chapter 2 provides some background technical information on related subjects such as security and privacy, electronic health data, cloud computing, cryptography and Attribute Based Encryption (ABE). It also provides some background on Attribute Based Encryption together with a review of literature outlining the work that has been done in the area of ABE in order to improve the security and performance original schemes while adding features such as revocation. A review of systems that have been built based on ABE for the handling of health related data is also included. Chapter 3 contains details about the proposed solution including the system and security models together with a detailed description of the entire protocol. Chapter 4 contains a discussion of the results from the implementation and Chapter 5 concludes the thesis and provides information on possible future improvements.

Chapter 2

Background and Literature Review

2.1 Security and Privacy

2.1.1 Security

Security according to the National Institute of Standards and Technology (NIST) [11] could be defined as any condition that leads to the creation and maintenance of defensive measures to ensure than an information technology infrastructure continues to perform it basic or critical functions irrespective of the risks posed by the threats to its normal operation.

The major objectives that need to be considered when analyzing the security of any system are Confidentiality, Integrity and Availability. These objectives are to be met completely for any system to be considered secure.

Confidentiality [11] assures that only those entities in the system that are authorized

have access to data that is either being stored, processed or transferred in the system.

Integrity [11] relates to the verification of the authenticity of data. This means ensuring that data has not been manipulated in any form either while in transit or while in storage. This ensures that any unauthorized manipulation of data in the form of addition, deletion or substitution is detected.

Availability [11] can simply be described as a measure of the level of accessibility and usability of a particular system upon request by an authorized user. This means that the system should be able to at all times carry out the various functions in order to meet the demands of its users. This also covers the ability of the infrastructure to remain functional even when some individual.

2.1.2 Privacy

Privacy is commonly equated with the concept of confidentiality although they are both distinct. While confidentiality is mainly concerned with ensuring that only users who are authorized have access to data which is being stored, processed or transferred within a system. Privacy, on the other hand, is concerned with ensuring that users have more control over the collection, use and storage of information that is related to them. Therefore, while maintaining the confidentiality of a system aids in preserving privacy, confidentiality does not completely ensure privacy as an authorized user may abuse that privilege by violating the privacy of user information [12].

The range of what is considered private information significantly varies in scope depending on the application area. For instance, in health sector, private information can be regarded as any oral or written information that meets any of the following

criteria: relates to the health of the individual, including their family history; relates to health care provision, including the source of care; constitutes as service for individuals who require long term care; relates to payment for health care. More importantly, private information is any information that can be used to identify an individual, either alone or when related to another piece of available information [13].

The privacy of data that is stored in the cloud faces multiple challenges as a result of the different ways in which the data are stored or processed on a machine that is usually owned by a different organization, the CSP. The major issues that exist in this area of privacy relate to trust as users are not completely certain that: their data is not being used for other purposes other than that for which it was collected; that data is destroyed properly in the end; privacy breaches have occurred which may have exposed their information; their information is retained even after they have stopped using a particular service [12].

2.2 Electronic Health Data

There are three broad classes in relation to the electronic collection and storage of health information and according to Canada Health Infoway, established for accelerating the adoption and use of digital health solutions across Canada, can be defined as:

- Electronic Health Record (EHR) [14] - these records usually contain information about an individual's health and their health care history. Typical information contained in these records include lab results, medication profiles, clinical reports, and diagnostic images. The EHR is made available electronically to

authorized health care institutions

- Electronic Medical Record (EMR) [14] - this refers to the digital form of the information acquired during an individual's visit to a health institution. This allows the doctor at the facility to gain access to information about the individual, including potentially information stored in the EHR.

- Personal Health Record (PHR) [14] - this is simply a compete of partial record containing information about an individual's health and usually in their custody. The health care institution has no control of this and it is managed by the individual.

The focus of this work is on electronic health records which are typically shared among multiple parties and under the control of the medical institution. Important point so needs to be reverified with source cited also and also additional sources

2.3 Cloud Computing

Cloud computing as defined by the National Institute of Standards and Technology (NIST) [15], is "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." Cloud computing offers considerable advantages to both government and private organizations, which has led to its growth and world wide acceptance in recent years. Some of the advantages offered by the cloud include: easy and fast deployment of IT systems; reduction in the cost of installation and maintenance of infrastructure; easy accessibility; improved flexibility

of systems; and a heavy reduction in the responsibilities of the user as most of the traditional tasks will be handled by the provider of the cloud based service, the CSP.

The different service models for cloud computing as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). These delivery models are distinct based on what services the CSP provides and the amount of responsibility that falls on the user in terms of control and management of resources. The IaaS model gives users more responsibilities as they have control over their operating systems, storage and applications which have been deployed while the SaaS offers the least amount of responsibilities which are limited to some application configuration settings. More detailed information about the different service models can be found at [15].

The deployment models available in cloud computing are the private, community, public and hybrid cloud models. These models are based on the number of parties that share the available deployed infrastructure. The private cloud is typically setup for use for a single organization while the community and public cloud models usually involve multiple parties with the former involving parties that share similar interest and requirements, while the latter is typically provisioned and available for use by the general public. The hybrid cloud model is basically a combination of the any of the other models and is typically a combination of the private and public models with the aim of benefiting from the strengths of the models while eliminating individual model weaknesses. More detailed information about the different deployment models can be found at [15].

The cloud computing deployment model this thesis considers is the public cloud model as this is the model mostly used or a hybrid model potentially involving a private and

public cloud model focusing on the vulnerabilities of the public facing infrastructure. Also, users of the private model have more control over their infrastructure and are able, to a certain degree, to ensure that the security and privacy of stored data is assured.

2.4 Cryptography

Cryptography [11] is the field of study which represents the principles, means and methods used for transforming data in order to hide their original content and prevent unauthorized use or modification. This typically involves the study of several mathematical techniques. Cryptography can be broadly divided into secret key and public key cryptography also known as symmetric and asymmetric schemes.

2.4.1 Secret Key Cryptography (Symmetric)

This type of cryptographic systems involve the use of a single secret key which is usually agreed upon by both parties who want to keep their communication secret. This secret key is used to encrypt the original message typically described as the plaintext (i.e encode the plaintext into a ciphertext that cannot be read by a party without the secret key). The receiving party if authorized and in possession of the secret key is able to decrypt the ciphertext and gain access to the original message. Examples of secret key schemes include the Ciphers (Caesar, Monoalphabetic and Polyalphabetic cyphers), Data Encryption Standard (DES) and Advanced Encryption Standard.

2.4.2 Public Key Cryptography (Asymmetric)

This type of cryptographic systems was developed as a result of the challenges in secret key cryptography which include the problem of key management and lack of secure channel for users to exchange keys. Public key cryptography involves the use of two separate keys, a public and private key, which are used to perform complementary operations such as encryption and decryption or signature generation and verification. Examples of public key schemes include the Diffie–Hellman key exchange protocol, RSA, Elgamal and Elliptic Curve Cryptography.

2.4.3 Cryptographic Adversary Models

Attacks on cryptographic systems typically carried out by an adversary normally to either recover the plaintext from the ciphertext or recover the secret key used by the system can be classified into four broad categories.

- Ciphertext Only Attack [16] - A ciphertext only attack is a class of attacks in which the adversary only has access to some ciphertext without any knowledge of the corresponding plaintext. This is the weakest type of attack because the adversary has the least amount of information to work with and any encryption scheme vulnerable to this class of attack is considered to be completely insecure.

- Known Ciphertext Attack [16] - A known plaintext attack is a class of attacks in which the adversary has access to some plaintext and ciphertext pairs. The adversary is unable to create more pairs and in only able to gain access to these by eavesdropping on the comunication channel between parties. These types of attacks are only marginally more difficult to mount.

- Chosen Plaintext Attack (CPA) [16] - A chosen plaintext attack is a class of attacks where the adversary is able to select the plaintext and request for the corresponding ciphertexts. This is typically done through the use of a black box system, typically called an oracle, that is able to produce the corresponding ciphertext when given any plaintext without revealing the key or any information about the plaintext of the original ciphertext that the adversary is trying to decrypt. A variation of this is the adaptive chosen plaintext attack where the adversary chooses the new plaintext based on the ciphertext received for earlier submitted plaintexts.

- Chosen Ciphertext Attack (CCA) [16] - A chosen ciphertext attack is a class of attacks where the adversary selects any ciphertext and requests for the corresponding plaintext. This is the direct opposite of the chosen plaintext attack class. This class of attacks are considered to be the strongest model of attacks when classifying encryption schemes based on their level of resistance. An adaptive chosen ciphertext attack just like the adaptive version of CPA, involves the adversary deciding on what ciphertext to submit based on the plaintext received for earlier requests.

Note that some of the attack types above are mutually exclusive (for instance, an attack cannot be both chosen plaintext and known plaintext). And also the chosen plaintext/ciphertext attacks are somewhat exclusive to the modern era of cryptography.

2.5 Technical Preliminaries

2.5.1 Bilinear Maps

The concept of bilinear maps, or pairings, are the foundation of pairing based cryptography which allowed of the creating of cryptosystems with a great variety of functionalities. Cyclic groups with efficiently computable bilinear maps form the basis of bilinear maps.

There are two general types of bilinear maps, namely:

1. **Symmetric Pairing [17]** - In this type of pairing we have two cyclic groups G, G_T of prime order p with g as the a generator of G. The efficiently computable bilinear map in this case is represented as

$$e\colon G \times G = G_T$$

2. **Assymmetric Pairing [17]** - In this type of pairing we have three cyclic groups G_1, G_2, G_T with G_1 and G_T of order p and G_2 a group with each element having an order dividing pG. The efficiently computable bilinear map in this case is represented as

$$e\colon G_1 \times G_2 = G_T$$

The type of pairing used in this work is symmetric. Below is a definition of bilinear maps as well as the properties that make it efficient for use in Attribute Based Encryption.

Definition 2.5.1. (Bilinear Maps [17]) Let G, G_T be two cyclic groups (multiplica-

tive or additive) of prime order p. Let g be a generator for G and $e\colon G \times G = G_T$. The bilinear map e has the following properties:

1. Bilinearity: for all $u, v \in Ganda, b \in Z_p$, we have $e(u^a, v^b) = e(u, v)^{ab}$.

2. Non-degeneracy: $e(g, g) \neq 1$

2.5.2 Access Structures

The definitions of access structures and linear secret sharing schemes used in this thesis have been adapted from [18].

Definition 2.5.2. (Access Structures [18]) *Let $\{P_1, \ldots, P_n\}$ be a set of parties. A collection $\mathbb{A} \subseteq 2^{\{P_1,\ldots,P_n\}}$ is monotone if $\forall B, C\colon$ if $B \in \mathbb{A}$ and $B \subseteq C$, then $C \in \mathbb{A}$. An access structure, i.e monotone is a collection $\mathbb{A}$ of non-empty subsets of $\{P_1, \ldots, P_n\}$ i.e., $\mathbb{A} \subseteq 2^{\{P_1,\ldots,P_n\}} \setminus \{\}$. The sets in $\mathbb{A}$ are called the authorized sets, and the sets not in $\mathbb{A}$ are called the unauthorized sets.*

In ABE, attributes play the role of parties in the access structure and the scheme in this thesis only considers access structures that are monotone.

Access policies based on monotone access structures could be represented as either a Linear Secret Sharing Scheme (LSSS) Matrix or with the use of monotonic boolean formulas which could be represented as an access tree in which the core nodes are used to represent the AND and OR gates with the attributes represented by the leaf nodes.

Access policies based on monotone access structures could be represented in two different ways for ABE schemes. The two widely used methods are:

2.5.3 Secret Sharing Schemes

Secret sharing schemes which was first created by Shamir in [19] allows for the division of data among multiple parties in such a way that the original data can only be reconstructed if a party is in possession of at least a fixed number of division, usually the threshold, and possession of a number of pieces less than the threshold reveals no information about the original data. Other earlier works in secret sharing include works by Barkley [20], Benaloh [21] and Ito, Saito and Nishizeki [22]. LSSS are secret sharing schemes in which the reconstruction of the original secret is done using a linear function of the available pieces [18].

2.5.4 Linear Secret Sharing Schemes

Definition 2.5.3. (Linear Secret Sharing Schemes [18] [23]) A secret sharing scheme $\prod$ over a set of parties $\mathcal{P}$ is called linear (over Z_p) if

1. The shared for each party form a vector over Z_p.

2. There exists a matrix M with ℓ rows and n columns called the share-generating matrix for $\prod$. For all $i = 1, /ldots, \ell$, with M_i representing the i'th row of M. The function ρ is defined as the party labeling row i as $\rho(i)$. Consider a column vector $\vec{v} = (s, r_2, \ldots, r_n)$ where $s \in Z_p$ is the secret to be shared and $r_2, \ldots, r_n \in Z_p$ are chosen randomly, then M_v is can be described as the vector ℓ shares of the secret s according to $\prod$. The share $(M_v)_i$ belongs to party $\rho(i)$.

The linear reconstruction property as described in [18] shows that suppose that $\prod$ is an LSSS for an access structure $\mathbb{A}$. Let $S \in \mathbb{A}$ be any authorized set, and let

$I \subset 1, /ldots, \ell$ be defined as $I = i : \rho(i) \in S$. Then there exits constants $w_i \in Z_{p_{i \in I}}$ such that, if ρ_i are shares of the secret s according to $\prod$, then $\sum_{i \in I} w_i \lambda_i = s$.

Note that access structures represented as boolean formulas which are typically represented by binary trees can be converted into a LSSS form using the techniques described in [24] with the number of rows in the corresponding matrix equal to the number of leaf nodes in the access tree.

2.5.5 Cryptographic Hardness Problems

Let G be a cyclic group of prime order p. Let g be a generator for G represented as $G = \langle g \rangle$ and let $x, y, z \in G$. The different complexity assumptions used to show the security of the different pairing based schemes have their foundation in the following core hardness problems [17]:

- **Discrete Log DLog) Problem** - Given g and g^x, compute x.
- **Computational Diffie-Hellman (CDH) Problem** - Given g, g^x, and g^y, compute g^{xy}.
- **Decisional Diffie-Hellman (DDH) Problem** - Given g, g^x, g^y and g^z , determine if $xy = z$.

2.6 Attribute Based Encryption (ABE)

ABE is a pairing based cryptogaphic scheme that was developed based on Identity Based Encryption(IBE) which was originally proposed by Shamir in 1984 [25]. Shamir

proposed a scheme which allowed for the encryption and decryption of information between two different users without the need for any exchange of keys between both parties. His proposal assumed the existence of a trusted key generation service similar to Certificate Authorities (CA) which were responsible for registration of users as they join a network and also for subsequent verification of their identity. Personal information unique to several users, such as their address, email address or a combination of this information, was used as the public key in the system. This allowed for the encryption of data meant for UserB by UserA using the email address of UserB, e.g "userB@gmail.com". UserB on receiving this encrypted data would then contact the key generation service and, after successful authentication, receives a secret key granting him access to the original datae. The scheme proposed by Shamir was further developed and the first practical and secure IBE scheme was presented by Boneh and Franklin in [26], who developed a fully functional IBE scheme which made use of groups for which there existed an efficiently computable bilinear map such as the Weil pairing.

Sahai and Waters in [10] were able to develop a new scheme that improved on the existing IBE schemes by creating a system in which the user identity is viewed as a set of descriptive attributes, allowing a user to encrypt data for all users who have a certain set of attributes. Decryption in this case is only permitted if the identity of a user, and the identity for which the ciphertext was encrypted, were close enough based on their individual attributes. It is in this work that the notion of Attribute Based Encryption is first mentioned.

Goyal et al. in [27] developed an ABE scheme that was more expressive than the original ABE scheme proposed by Sahai and Waters [10]. In their scheme, termed Key-Policy Attribute-Based Encryption (KP-ABE), each ciphertext created by the

user contains a set of descriptive attributes. Secret keys of individual users are associated with an access structure which specifies the attributes that need to be contained in a ciphertext for successful decryption. The access tree structure could be made up of interior nodes that consist of AND and OR gates with the leaves containing different attributes. For example, if UserA's key in KP-ABE contains "C AND D" as the access policy, the only ciphertexts he should be able to decrypt are those that contains both attributes C and D. A ciphertext with only attribute C or D could not be decrypted by UserA as the requirements for access would not be satisfied. The keys generated for users in this scheme are also collusion resistant just like the original scheme, meaning that no two users with different attributes could combine their keys to create an overlap of attributes that would give them the ability to decrypt files which they would not normally be able to decrypt.

The authors in [27] mentioned a variant to the KP-ABE scheme known as the Ciphertext-Policy Attribute-Based Encryption (CP-ABE) scheme which they left as an open problem that was solved by [28]. In CP-ABE, the ciphertexts are associated with the access policy while the user keys contain a set of descriptive attributes. This would mean that, for a key to decrypt a particular ciphertext, its attributes need to match the access structure of the access policy of the ciphertext. This scheme, unlike the KP-ABE scheme, gives the user encrypting data more control as the user is able to control who can have access to data being encrypted by making sure the access policy in the ciphertext specifies what attributes need to be possessed for access to be granted.

The four basic algorithms of any ABE based system includes the following:

1. Setup - The setup algorithm is responsible for the selection of the bilinear group

and the definition of a bilinear map that has the properties of bilinearity, computability and non-degeneracy. The setup algorithm takes as its input the security parameter which specifies the size of the attribute set and generates a public key (PK) and a master key (MK) as output.

2. Keygen - The keygen algorithm takes as its input two parameters, the MK generated during setup and the set of attributes that the user possesses, and generates a secret key (SK) for the user in CP-ABE. The input for KP-ABE includes the MK, PK and an access structure and outputs SK.

3. Encryption - The encryption algorithm takes as its input PK, a message M, and an access structure for CP-ABE schemes and produces a ciphertext C. It takes as input PK, M and a set of attributes and produces a ciphertext C for KP-ABE schemes.

4. Decryption - The decryption algorithm takes as input PK, C and SK and, if the attributes of either the ciphertext or secret key satisfies the access structure of the other, depending on whether the scheme is a CP-ABE or KP-ABE scheme, decrypts C and outputs M.

Both variants of ABE allows for delegation which would allow a user with a secret key for set of attributes (CP-ABE) or containing an access structure (KP-ABE) to derive a new secret key containing a set of attributes that is a subset of the attribute set of the original key or an access structure that is more restrictive than the access structure contained in the original key.

2.7 Revocation

A useful feature with ABE based systems is the ability for user revocation. This is a challenge as multiple users may share a similar attribute which could cause the revocation of one user to affect other users who share a similar attribute. Furthermore it is important that user revocation be flexible and occur at different granular levels which means that revocation could involve removal of a user completely or a partial reduction of a user's access, based on their attributes. The addition of an expiry date to the generated key has been proposed by initial ABE based systems [28] but does not offer an effective means for the revocation of user attributes.

Yu et al. in [29] proposed a scheme that enables secure, scalable and fine grained data access to a cloud based system with a great reduction in the computation overhead by delegating most of the computation intensive tasks to the cloud servers while ensuring the security and privacy of user data through the combination of KP-ABE, proxy re-encryption (PRE) and lazy re-encryption (LRE). In their scheme, the attributes that are assigned to the ciphertexts are all assigned a unique ID which serves as the version number that is stored in a list maintained by the cloud servers, together with the PRE keys used. The cloud servers also maintain a list of all the existing users in the system who are currently authorized to have access to the different stored data. Data files are encrypted using symmetric encryption with the decryption keys encrypted using ABE and appended to the encrypted data file together with a unique file ID. PRE enables the use of a proxy to convert a ciphertext which has been encrypted using the public key of a particular user into another ciphertext that can be decrypted using the private key of a different user, without revealing the contents of the underlying file. In order to revoke a user, the scheme determines the least amount of attributes that

need to be updated to prevent a user from having access and redefines the public and master keys for those attributes while generating the corresponding PRE keys. The revoked user's ID, the attribute set, the PRE keys and the new public key parameters are sent to the cloud servers. The cloud servers then remove the revoked user from the user list, store the new public key parameters and then updates its list of attributes together with the PRE keys used.

Yu et al. in [30] have applied the concept of proxy re-encryption with CP-ABE in order to enable revocation. Liang et al. in [31] have proposed the system Ciphertext Policy Attribute Encryption with Revocation (CP-ABE-R), which makes use of linear secret sharing and binary tree techniques to enable effective revocation of users with the aid of a unique identifier assigned to each user in the system and which is not needed for encryption and decryption. Cheng et al. have proposed a scheme [32] that enables effective revocation in CP-ABE by dividing the original data into multiple parts which they term *slices*, before they are stored in the cloud which allows for revocation by the re-encryption of only one slice. The data is encrypted with a symmetric key and then split into multiple parts using a secret sharing scheme. In the case of the secret sharing scheme applied here, the number of parts that are needed to reconstruct the original file is equal to the number of distinct parts. A particular slice of data is chosen as the dynamic data and it is this slice that is constantly re-encrypted to enable revocation while the static data remains the same. This reduces the computational and storage overhead while ensuring that the security of the system is not compromised.

2.8 Multi-Authority Schemes

The first Multi-Authority Attribute Based Encryption (MA-ABE) scheme was proposed by Chase in [33] [34] and was based on Key Policy Attribute Based Encryption (KP-ABE). In this scheme, there are multiple Attribute Authorities (AA) in addition to a Central Authority (CA) which are responsible for generating secret keys corresponding to the attributes which they handle. Users are assigned a Global Unique Identifier (GUID) which they use to request the shares of the system-wide Master Secret Key (MSK) handled by the different authorities. The GUID is used by the authorities to tie the shares to a particular user. The system includes a CA which is responsible for aiding users in decryption by ensuring that all the shares generated for a particular user by the different AAs sum up to the same MSK. The CA ensures this by assigning to each user a special value that cancels out all of the shares from the different authorities, providing the user with a function of the system wide MSK. In order to carry out its functions, the CA has knowledge of the MSK and as a result would also be able to decrypt any ciphertext in the system which is in contrast with the idea behind using multiple authorities, which is to distribute trust among several untrusted authorities. Also, in the original system, the users use their GUIDs to identify with the individual AAs, which means that several authorities could combine their information about a particular user and develop a profile based on the attributes that the user has acquired and be able to generate keys with the same level of access as the user.

Chase and Chow in [35] proposed a solution to the original MA-ABE problem which eliminated the use of a CA and also prevented the AAs from having the ability to combine the information about a user by allowing users to use pseudonyms for

interacting with the individual AAs in the system instead of the use of their respective GUID. This solution eliminates the CA by applying a set of Pseudo Random Functions (PRF), and having every pair of authorities in the system share a secret PRF seed which allows for a combination of all their individually generated shares. To enable users to communicate with the individual AAs using pseudonyms, the authors have developed a novel Anonymous Key Issuing Protocol. Additional details about the protocol and its functions can be found in [34] [35].

Lewko and Waters propose a MA-ABE [36] [24] solution in which the different authorities operate independently and do not have to share any common information with each other as in [33] e except for an initial set of common reference parameters. Their system has higher tolerance as the failure or corruption of authorities in the system will not have a direct impact on the operation of the fully functioning and uncorrupted authorities. Furthermore, in their solution, any party could become an authority by making available, to the other entities in the system, their verification key and their lift of managed attributes. This solution makes use of Linear Secret Sharing Schemes (LSSS) access structures and the authors show that boolean formulas could easily be transformed into LSSS structures using techniques found in [24]. Lewko and Waters have used the dual system proof technique to prove the security of their system.

Kan et al. in [37] have developed a solution called DAC-MACS (Data Access Control for Multi-Authority Cloud Storage) in order to make a more efficient CP-ABE based MA-ABE solution that takes advantage of the services available in the area of cloud computing and that is more suited to this domain. Their solution includes an efficient attribute revocation method that enables both forward and backward security. In addition to providing a means for attribute revocation, Kan et al.'s system [37] has

better efficiency than other similar solutions. Also, by using a decryption token for the decryption, they have been able to transfer the intensive computations over to the cloud server, thereby reducing the computational overhead on the side of the end user. A flaw in the system is the fact that the different AAs have knowledge of the GUID of the users which would give a revoked user the ability to derive the key update key that they could use to update their own keys by corrupting any AA, together with some non-revoked users.

Kan and Xiaohua [38] have developed a more effective MA-ABE solution based on CP-ABE. In their new system, the secret keys of the different users are not related to the key of the data owner and so users will only need to hold an individual secret key for an authority instead of multiple secret keys associated with multiple owners. This makes it more suitable for a multiowner setting as storage overhead for user keys is greatly reduced. They have also improved the revocation mechanism by modifying it to require that only ciphertexts associated with a revoked attribute be updated and by using a single update key for the update of both keys and ciphertexts.

2.9 Outsourcing

- main goal? (ref original paper [39])
- optimization? (ref subsequent papers)
- method (i.e split decryption into two phases [cite original paper i.e [39]]
- advantages (allow for low compute devices to be employed at the user end. why?)

- risk? (none - same level of security and privacy as the original ABE approach)

Give general info about outsourcing in ABE to provide background knowledge on approach used to provide outsourcing capabilities to ABE scheme in this paper

- Aim of outsourcing i.e splitting the decryption process for greater privacy and security while reducing the operations on the user end opening access to the use of devices with lower computational power.
- split into an elgamal ciphertext cite original outsourcing paper for more info
- other advantages of outsourcing from other papers that used a similar technique and what their approaches were also if necessary

2.10 ABE Based Health Care Systems

Ibraimi et al. proposed a new variant of CP-ABE [2] in order to be able to enforce the required levels of access controls in a multiple domain based system to ensure the security of personal health records (PHR). They have distributed the group of users who normally require access to PHRs into two domains: the professional domain, which consists of the health care providers; and the social domain, made up of family members, friends and possibly fellow patients. Their proposed variant of CP-ABE allows the encryption of health records with an access policy that is made up of attributes issued by two different trusted authorities: the trusted authority in charge of the professional domain and the trusted authority in charge of the social domain.

The authors in [3] have proposed the design of a patient controlled cloud based EHR infrastructure using CP-ABE. They have based their system on the assumptions: a

trusted authority (TA) exists that is responsible for the generation of keys for users and is able to store the public parameters and public keys of users in a public directory; each user is associated with a unique identifier and a set of attributes; and the cloud server used for storage is only trusted for the performance of storage operations. They have used a variant of CP-ABE, known as broadcast ciphertext-policy attribute-based encryption (bABE), which extends the traditional CP-ABE to enable revocation of users' keys. They have also provided the functionality of keyword search which allows users to search using a search term by providing a key which allows the cloud provider to perform search operations on the encrypted data without learning anything about the actual data contents.

Akinyele in [4], using ABE, has provided a detailed design and implementation of self-protecting EMR which allows EMR availability even when the providers are offline. Their system makes role- and content-based access control possible. For role-based access control, users are granted explicit access to collections of data related to their roles that match some specific criteria which the authors have termed content slices. These slices could be as specific as required by the system administrators and the content-based access is used to grant access to users such as contractors, who have no definite roles in the system but require access to records to carry out their functions. They have also implemented a policy engine as part of their design to evaluate new or updated EMRs in order to determine the policies that are to be used for encryption. The policy engine's final decision is based on either the set of policies specified by the administrator, the identity and nature of the EHR, the annotations attached to the EHR, or in some cases the textual content of the record. They have taken advantage of the XML-based EMR standards which include the Continue of Care Record (CCR) and the Continuity of Care Document (CCD) which allows their policy encryption

engine to parse each node in order to determine the appropriate access policy and subsequently the access control rule and content related attributes for which the document is to be encrypted. Users will need to be present at the initialization stage to have their mobile devices provisioned with the required decryption keys to be able to use the accompanying mobile application to access their data. CP-ABE is used to grant access to patients and health professionals using keys with fixed attributes related to their roles or responsibilities while users with no definite role are granted access through the use of KP-ABE by generating keys, which contain a specific policy that defines what data they can access and, in some cases, the time periods for which they can have access.

Barua et al. in [5] have proposed a scheme which they called Efficient and Secure Patient-centric Access Control (ESPAC), in which they have used CP-ABE to achieve patient-centric access control allowing different access privileges based on the roles of the data requester and assigning the corresponding attributes based on those privileges. They have constructed their access control policy by assigning attributes, based on the relationship between the patient and the requesting party, which is used to determine the privacy levels of the requesting party before attributes are assigned. Their system is made up of four main entities: the trusted authority (TA), the cloud service provider (CSP), the registered user, and the data-access requester. The scheme makes use of pseudo identity instead of unique identities to ensure privacy. The scheme enables message integrity checks, non-repudiation and source authentication through the use of signature verification. This scheme is able to ensure forward and backward secrecy.

Suhair et al. in [6] have proposed the design of a cloud based EHR system using CP-ABE to ensure security by using the credentials and attributes of the health

care providers as the universal attribute set. Their proposed architecture is made up mainly of three components: the EHR system hosted on the cloud; the participating healthcare providers; and the attribute authority (AA) which is in control of generating the secret keys of users which contain the appropriate attributes. The cloud is used for data storage and computation in their infrastructure. Encryption and decryption of the medical records are performed at the client end through the use of lightweight software. Suhair et al. have proposed the addition of an expiry date to the access policies used for encryption, or complete re-encryption with updated access policies, as a way to achieve revocation in order to avoid the communication overhead involved with the re-distribution of secret keys to authorized users. The use of a single AA presents a focal point of weakness for the security of the system and presents the key escrow problem.

Hupperich et al. in [7] propose an architecture that gives the patient control of the delegation of access to their EHRs, in line with the existing privacy laws. They have proposed a system that would allow patients to authorize the appropriate health care service providers to have access to their EHRs through a flexible channel that would not require the patient to be present. In order to eliminate the use of smart cards for access, and to enable health care providers to have access to EHRs their infrastructure only requires the use of the patient's smart card at the initial stage for the generation of a transaction code (TAC) which the patient can use to grant access by sending to authorized health care professionals. They have used ABE for encryption by using the patient's identity and a TAC that is specific to a particular medical record as the two main attributes for encryption and decryption. They have implemented emergency access by allowing the encryption of certain records using the attribute *"emergency"* without any TAC, with logging implemented to keep track of emergency access. The

authors have not mentioned how the system would handle revocation of users and have not implemented a secure means of transmission for the generated TACs which they have stated could be transferred via traditional means such as a phone or on paper.

The authors in [8] proposed the design for a secure interoperable cloud based service for private health records (PHR) which uses the Continuity of Care Document (CCD) for the storage and exchange of information and employs several security mechanisms using available open standards such as XACML, XML encryption, XML signature and XML key management specification. They have used CP-ABE to achieve patient controlled encryption, and the public key encryption with keyword search (PEKS) scheme to provide privacy-preserving keyword search. They have used the Secure Channel Free PEKS scheme which allows users to perform private searches over encrypted data for specific matching keywords without revealing the keywords or any partial matches to the server.

Li et al. in [9] proposed a framework titled Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute Based Encryption containing a suite of security mechanisms that aimed to solve the existing issues in cloud-based PHR storage systems which include eliminating the risk of privacy exposure, key management scalability, flexible access and effective revocation of existing users. Their work focuses on the multiple data owner scenario similar to our proposed architecture and thus they have divided the users in the system into two broad security domains similar to [2], which reduces the complexity of key management for data owners and users who require access, with the improvement being that in their scheme the public domain (PUD) is managed by multiple AAs. The personal domain is made up of users who are close to the data owners (i.e. patients) such as family members and

friends while the public domain is made up of the various professionals who require access to the patient's records such as doctors and pharmacists. In order to apply ABE to the personal domain, Li et al. have employed the Key Policy ABE with efficient revocation as proposed in [30] with the data owner fully responsible for handling this particular domain. The data owner generates keys for members of this domain with the access structure corresponding to their level of access and sends this keys to the corresponding users in order to grant them access. The authors have employed in the public domain the use of the MA-ABE proposed in [33] and improved in [35]. Since the MA-ABE scheme they adopted is essentially a KP-ABE scheme with multiple AAs, in which control of access lies with the AAs who generate the keys for the different attributes therefore taking away control from the data owner, Li et al. have made a slight modification to how this scheme is used in their system. In order to grant the data owner more control, the system requires that the key access policies and the general approach for specifying the ciphertext attributes be agreed upon in order to grant the users some level of control in specifying the access policy of the ciphertext from their end by choosing the right attributes. To improve security in the public domain, Li et al. have slightly modified the Multi-authority ABE (MA-ABE) scheme proposed by [35] to enable efficient user revocation by using the revocation technique proposed by [30] which was not a feature of the original scheme. Their system provides dynamic attribute and access policy modification together with on-demand user/attribute revocation, together with break glass access, in order to make records available for use under emergency situations.

Chapter 3

Secure Privacy Preserving Framework for Elctronic Health Records (EHRs)

3.1 Overview

- what the chapter will contain: description of the proposed framework for secure storage of data using a combination of attribute and symmetric-based encryption. details of implementation. Also provide info on the security model for the underlying attribute based encryption scheme. This would highlight the assumptions that was used to verifying the integrity of the proposed scheme. I will summarize the pros and cons of the proposed scheme.

- what are the good things about the model: its more secure than xyz, runs faster.

potential drawbacks: ... its own section?

- what the subsequent sections will contain to provide a clearer image of their content

3.2 Multiple Authority Ciphertext Policy Attribute Based Encryption with Outsourced Decryption

3.2.1 ABE Scheme Definition

Our ABE scheme has been adapted from the scheme developed in [cite oiriginal paper]. Using similar methods applied in [cite outsourcing paper], we have modified the scheme with the aim of outsourcing the bulk of the decryption algorithm operation to the CSP. This section contains a definition of the algorithms that make up our ABE scheme.

Setup Algorithms

The setup algorithms are run as a part of the initialization stage of the ABE scheme. These algorithms are responsible for the registration of the other entities in the system. This also provides them with the necessary information for verifying their identity when communicating with other entities in the system and other vital information they need in order to execute their own operations

- CASetup - The CASetup algorithm is the algorithm executed by the Certificate Authority in order to generate the Global Master Key (GMK) and Global Public Parameters (GPP) of the entire scheme. It takes as input the security parameter of the scheme. (look up impact of that security parameter...has to do with key length and curve size and stuff like that most likely). The CASetup algorithm also takes care of the registration of the users and attribute authorities (AAs) in the system by assigning to them their unique ids. It is also responsible for generating the public-secret key pairs together with the user certificates and their corresponding verification keys. It makes the GPP available to all registered entities in the system and sends an unmatched public-private key pair for all users to the AA.

- AASetup - The AASetup algorithm is run by the individual Attribute Authorities (AAs) and generates the public-secret key pair for the attribute authority. It also generates a public-version key pair for each of the attributes under the control of the AA that is provided as input in the AA universal attribute set.

Key Generation Algorithm

- SKGen - The SKGen algorithm is run by the corresponding AAs in order to provide the user with the secret key for the attributes that they have been assigned. The AA takes as input the system GPP, the users Global Public Keys (GPKs), one of the user GSKs, the set of attributes and their corresponding public-version key pairs. It provides the user with a secret key that can be used for decryption.

Encryption Algorithm

- Encrypt - The Encrypt algorithm is run by the data owner in order to encrypt the corresponding piece of data. This algorithm takes as input the system GPP, the public keys (PKs) for the AAs involved in the encryption - as they are in control of the attributes that make up the policy under which data is to be encrypted - of the data, and an access policy. It provides as output a ciphertext (CT) which implicitly contains the corresponding access policy.

Decryption Algorithms

- CTransform - The CTranform algorithm is responsible for the transformation of the CT into an El Gamal style ciphertext that can be easily decrypted by the end user. This algorithm is executed by the Cloud Service Provider (CSP) and takes as input the ciphertext, the secret keys of the attributes in the user's possession and the user's GPK. It produces as output a partially decrypted ciphertext (CT') is the user's secret keys meet the policy under which the data was encrypted, otherwise it outputs an error.

- Decrypt - The Decrypt algorithm is run by the end user and takes as input the partially decrypted ciphertext CT' and the user's global secret key which is not shared with any other entity in the system and produces the original data as output for the user in its plain form.

Revocation Algorithms

- UKeyGen - The UKeyGen algorithm is run by the AA under whose control the revoked attribute falls. The algorithm takes as input the SK of the AA, the revoked attribute and the current version key for the revoked attribute. It produces as output an updated version key for the revoked attribute. It produces an update key to be used to update the corresponding secret keys of users who possess the revoked attributes and whose access is not being revoked. It also produces an update key for the affected ciphertexts whose access policies contain the revoked attribute to allow for users who still maintain access to be able to decrypt while denying revoked users further access.

- SKUpdate - The SKUpdate algorithm is run by each non revoked user in the system and takes as input the corresponding secret key acquired from the AA that control the affected attribute together with the update key for user secret keys generated by the UKeyGen algorithm. It outputs a new secret key for the user.

- CTUpdate - The CTUpdate algorithm is run the CSP and takes as input the affected ciphertexts and the update key generated by the UKeyGen algorithm for ciphert update. it outputs a new ciphertext that can only be decrypted by users with a current version of the corresponding revoked attribute.

Security Model/Security Game + Assumptions....to be clarified and determined

- DDH Problem? - High level proofs e.g correctness, etc

- Security game? N/B Based on security games and also has to do with interactions to verify security of scheme between an attacker and a challenger. More details from reference materials

- What metrics from a security point of view is the system based on? CPA vs CCA

- What assumption(s) is from the point of view of security is the system to be analyzed by (to be adopted from original paper + possible outsourcing papers also and the basic correctness and the completeness high level assessments)

3.2.2 ABE Scheme Construction

A. CA Setup - This procedure is run by the CA to setup the system

$$CASetup(1^{\lambda}) \longrightarrow GMK, GPP$$

where 1^{λ} is the security parameter, GMK is the Global Master Key, and GPP represents the Global Public Parameters.

The CA chooses two multiplicative groups G, G_T with the same prime order p and a bilinear map $e\colon G \times G \to G_T$ and a hash function $H\colon \{0,1\} \to G$. The CA chooses two random numbers $a, b \in Z_p$

The GMK is set as (a, b) and the GPP is set to (g, g^a, g^b, H)

(i) User Registration

This is executed for all users in the system by the CA. Each user is assigned a globally unique uid and for each uid, the CA generates two random numbers $u_{uid}, u'_{uid} \in Z_p$ as its global secret keys

$$GSK_{uid} = u_{uid}, GSK'_{uid} = u'_{uid}$$

The global public keys for each user is generated as

$$GPK_{uid} = g^{u_{uid}}, GPK'_{uid} = g^{\frac{1}{u'_{uid}}}$$

The CA generates a $Certificateuid$ for each user uid and send (GPK_{uid}, GSK'_{uid}, $Cerificate(uid)$) to the user.

(ii) Attribute Authority (AA) Registration

The CA assigns a globally unique authority identity aid to each AA and sends (GPK'_{uid}, GSK_{uid}) for all registered users to AA_{aid}. The CA also sends the verification key vk_{CA} to AA_{aid} for verifying the $Certificate(uid)$ assigned to each user.

B. AA Setup

Let X_{aid} be the set of all attributes managed by AA_{aid}. The AA chooses three random numbers $\alpha_{aid}, \beta_{aid}, \gamma_{aid} \in Z_p$ as its secret key.

$$SK_{aid} = (\alpha_{aid}, \beta_{aid}, \gamma_{aid})$$
$$PK_{aid} = (e(g,g)^{\alpha_{aid}}, g^{\beta_{aid}}, g^{\frac{1}{\beta_{aid}}})$$

For each attribute $x_{aid} \in X_{aid}$, AA_{aid} generates a public attribute key as

$PK_{x_{aid}} = (PK_{1,x_{aid}} = H(x_{aid})^{v_{x_{aid}}}, PK_{2,x_{aid}} = H(x_{aid})^{v_{x_{aid}}\gamma_{aid}})$ where $v_{x_{aid}}$ is the version key of attribute x_{aid} i.e $VK_{x_{aid}} = v_{x_{aid}}$.

C. Secret Key Generation

The Attribute Authority with AA_{aid} assigns a set of attributes $S_{uid,aid}$ to user with *uid* after authentication and certificate verification. The AA chooses a random number $t_{uid,aid} \in Z_p$ and computes a secret key for the user as

$$SK_{uid,aid} = (K_{uid,aid} = g^{\frac{\alpha_{aid}}{v_{uid}}} g^{au_{uid}} g^{bt_{uid,aid}}, K'_{uid,aid} = g^{t_{uid,aid}},$$
$$\forall x_{aid} \in S_{uid,aid} \colon K_{x_{aid},uid} = g^{t_{uid,aid}\beta_{aid}} H(x_{aid})^{v_{x_{aid}}\beta_{aid}(u_{uid}+\gamma_{aid})})$$

If the user *uid* does not hold any attributes from AA_{aid}, the user secret key $SK_{uid,aid}$ only contains $K_{uid,aid}$.

D. Encryption

Data is divided into components based on the level of granularity required for access control i.e $m = \{m_i, \cdots, m_n\}$. Data components are encrypted using symmetric encryption keys $\kappa = \{\kappa_i, \cdots, \kappa_n\}$. An access structure (M_k, ρ) is defined for each content key $\kappa_i (i = 1, \cdots, n)$ and encrypted using the ABE scheme to produce the corresponding ciphertext. Let M be an ℓ x n matrix, where ℓ denotes the total number of all the attributes and the function ρ associates rows of M to attributes. The function ρ is not required to be injective which allows for an attribute to be associated with more than one row of M.

To encrypt the content key κ_i, the algorithm chooses a random element $s \in Z_p$ which is used as the random encryption exponent. It then selects a random vector $\vec{v} = (s, y_2, \ldots, y_n) \in Z_p$ where $y_2, \ldots, y_n$ are used to share the encryption

exponent s. It then computes $\forall 1 \leq i \leq \ell : \lambda_i = \vec{v}.M_i$ where M_i is the vector corresponding to the i-th row of M. It then randomly selects $r_1, r_2, \ldots, r_\ell$ and computes the ciphertext as

$$CT_{K_i} = \Bigg(C = K_i \cdot \Big(\prod_{aid_k \in I_A} PK_{aid_k}\Big)^s, C' = g^s, C'' = g^{bs},$$
$$\forall 1 \leq i \leq \ell, \rho(i) \in X_{aid_k} : C_i = g^{a\lambda_i \cdot \left(PK_{i,\rho(i)}\right)^{-r_i}}, C_i' = g^{r_i}, D_i = g^{\frac{r_i}{\beta_{aid_k}}}, D_i' = \Big(PK_{2,\rho_{(i)}}\Big)^{r_i}\Bigg)$$

Then the encrypted data is uploaded to the cloud server by the owner.

E. Decryption

(i) Phase 1 - Ciphertext Transformation (Partial Decryption)

This is the first phase of the data encryption component that involves the transformation of the ciphertext into an El Gamal style ciphertext referred to in this framework as a token through partial decryption while the integrity of the original message is preserved.

$$CT' = \prod_{aid_k \in I_A} \frac{e\Big(C', K_{uid,aid}\Big) e\Big(C'', K'_{uid,aid}\Big)^{-1}}{\prod_{i \in I_{aid_k}} \Big(e\Big(C_i, GPK_{uid}\Big) e\Big(D_i, K_{\rho(i),uid}\Big) e\Big(C_i', K^{-1}_{uid,aid_k}\Big) e\Big(g, D_i'\Big)^{-1}\Big)^{\omega_i n_A}}$$

N/B -

Look up the description of the process for the generation of the secret share using the shares represented by the attributes that are contained in the key. This technique has something to do with matrices. For potentially more details look up the papers by brent waters or allison lewko? for additional

information in a more summarized format. For more intense details look at the PhD thesis of ben lynn which is the main source

ω_i - constant chosen for the reconstruction of encryption exponent s using shares λ_i

n_A - number of AAs involved in the ciphertext

$\rho(i)$ - mapping of each row of access structure to attribute i

$$CT' = \prod_{aid_k \in I_A} e\Big(g, g\Big)^{\frac{s\alpha_{aid}}{u'_{uid}}}$$

(ii) Phase 2 - Data Decryption

The user does an exponentiation on the token to get the blinding element (BE) [rephrase this later] for decryption.

$$BE = CT'^{u'_{uid}} = \prod_{aid_k \in I_A} e\Big(g, g\Big)^{s\alpha_{aid}}$$

Remember the C element of the ciphertext $= K_i \cdot \big(\prod_{aid_k \in I_A} PK_{aid_k}\big)^s$ where $PK_{aid_k} = e\big(g, g\big)^{\alpha_{aid}}$. Therefore the original message which in this case is the symmetric key is computed as

$$K = \frac{C}{BE}$$

F. User Revocation

User revocation which can also be referred to as attribute revocation as used in some of the literature on the subject should achieve two critical criteria. The revoked user should not be able to decrypt new ciphertexts which have been encrypted using the public attribute keys of attributes that the user was previously granted private keys for. This is referred to as Backward Security. Any new user who has the right attributes should be able to decrypt the original ciphertexts that were encrypted using those public parameters. This is referred to a Forward Security. The revocation algorithm used has been gotten from the scheme by Yang et. al in [38].

to be formatted i.e more info added after clarification

math symbols to be used in this section

(i) Update Key Generation

In order to revoke an attribute $\tilde{x}_{aid'}$ from a particular user, the attribute authority (AA) responsible $AA_{aid'}$ runs the key update algorithm taking as input the secret key $SK_{aid'}$ of the attribute authority, the revoked attribute $\tilde{x}_{aid'}$, and its current version key $VK_{\tilde{x}_{aid'}}$. The algorithm generates a new version key $VK'_{\tilde{x}_{aid'}} = v'_{\tilde{x}_{aid'}} (v'_{\tilde{x}_{aid'}} \neq v_{\tilde{x}_{aid'}})$ for the revoked attribute $\tilde{x}_{ai'd}$.

The $AA_{aid'}$ then generates a unique update key $UK_{s,\tilde{x}_{aid'},uid}$ for secret key

update by each non-revoked user uid as

$$UK_{s,\tilde{x}_{aid'},uid} = H(\tilde{x}_{aid'})^{\beta_{aid'}(v'_{\tilde{x}_{aid'}} - v_{\tilde{x}_{aid'}})(u_{uid}+\gamma_{aid'})}$$

and generates the update key $UK_{c,\tilde{x}_{aid'},uid}$ for ciphertext update as

$$UK_{c,\tilde{x}_{aid'},uid} = \left(UK_{1,\tilde{x}_{aid'}} = \frac{v'_{\tilde{x}_{aid'}}}{v_{\tilde{x}_{aid'}}}, UK_{2,\tilde{x}_{aid'}} = \frac{v_{\tilde{x}_{aid'}} - v'_{\tilde{x}_{aid'}}}{v_{\tilde{x}_{aid'}}\gamma_{aid'}}\right)$$

The $AA_{aid'}$ sends the $UK_{s,\tilde{x}_{aid'},uid}$ to the non-revoked user uid and sends $UK_{c,\tilde{x}_{aid'},uid}$ to the cloud server.

The $AA_{aid'}$ then updates the public attribute key of the revoked attribute $\tilde{x}_{aid'}$ as

$$\widetilde{PK}_{\tilde{x}_{aid'}} = (PK_{\tilde{x}_{aid'}})^{UK_{1,\tilde{x}_{aid'}}}$$

(ii) Secret Key Update

This process is run by the non revoked users in the system. The algorithm produces an updated secret key using the update key $UK_{s,\tilde{x}_{aid'},uid}$ for the individual users as

$$\begin{aligned}\widetilde{SK}_{uid,aid'} = \Big(&\tilde{K}_{uid,aid'} = K_{uid,aid'}, \tilde{K'}_{uid,aid'} = K'_{uid,aid'},\\ &\tilde{K}_{\tilde{x}_{aid'},uid} = K_{\tilde{x}_{aid'},uid}.UK_{s,\tilde{x}_{aid'},uid},\\ &\forall x_{aid'} \in S_{uid,aid'} \backslash \{\tilde{x}_{aid'}\} \colon \tilde{K}_{\tilde{x}_{aid'},uid} = K_{\tilde{x}_{aid'},uid}\Big)\end{aligned}$$

(iii) Ciphertext Update

$$\widetilde{CT} = \Big(\tilde{C} = C, \tilde{C}' = C', \tilde{C}'' = C'',$$
$$\forall 1 \leq i \leq \ell : \tilde{C'}_i = C'_i, \tilde{D}_i = D_i,$$
$$if \rho(i) = \tilde{x}_{aid'} : \tilde{C'}_i = C_i.(D'_i)^{UK_{2,\tilde{x}_{aid'}}}, \tilde{D'}_i = (D'_i)^{UK_{1,\tilde{x}_{aid'}}},$$
$$if \rho(i) \neq \tilde{x}_{aid'} : \tilde{C}_i = C_i, \tilde{D'}_i = D'_i\Big)$$

3.3 Secure Privacy Preserving EHR Framework

- describe the different entities that make up the system and the roles they play (Update: rough draft prepared 10/01/2019)
- show system architecture diagram
- Process diagrams also

3.3.1 System Architecture/Model

System Entities

(1) Data Owner (DO) - The DO is responsible for the generation of data to be stored on the platform and retains ownership of data for the duration of its life cycle. The Data Owner also specifies what requirements are to be met for any user to have access by indicating what attributes they should possess though the policy used for data encryption. may need to indicate real life example....

(2) Data User (DU) - The DU represents entities that require access to stored data. They are required to possess the right attributes in order to gain access to stored data on the cloud.

N/B: It should be indicated in a real life situation that an individual or organization can play the role of both data owner and user in the system with the main difference being whether the posses ownership rights to the data in question for that specified scenario.

(3) Certificate Authority (CA) - The CA plays the central role of assisting with the verification of the identity of the different entities to enable them to securely communicate with one another. It does this by providing the corresponding certificates and public parameters used in the verification process.

(4) Attribute Authority (AA) - The AA is responsible for providing the appropriate public and private key parameters for the different attributes under their control to the appropriate owners and users for the encryption and decryption of data.

(5) Cloud Service Provider (CSP) - This entity provides the necessary storage facilities for data on the platform and also the processing capabilities for part of the operation related to the decryption of data.

- system architecture diagram

Figure 3.1 shows the complete architecture of the entire system framework, indicating the various entities and the different ways in which they interact with each other as part of the overall system.

vs

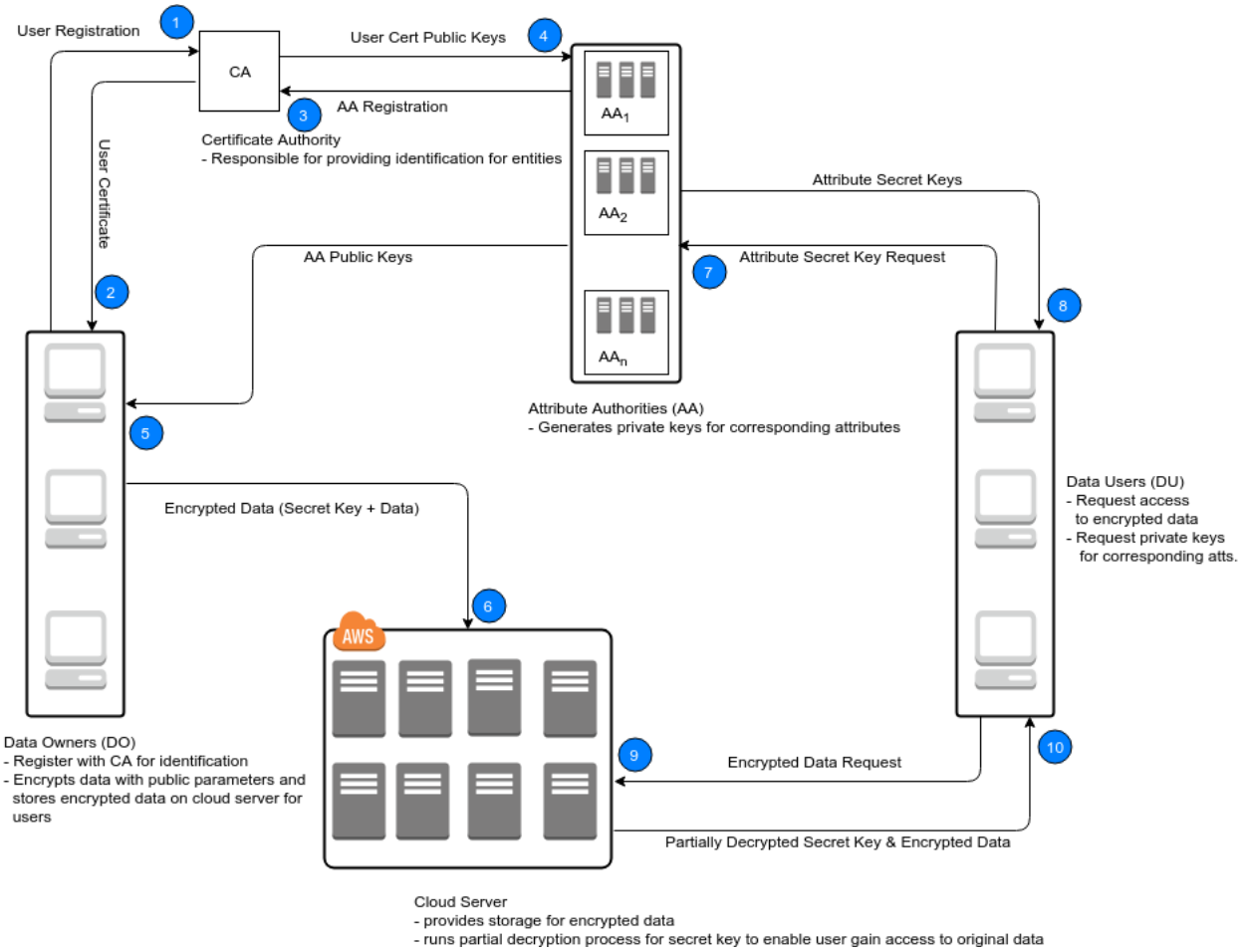

Figure 3.1: System Architecture Diagram

The figure shows a representation of the ABE framework representing the security and privacy of the system as a whole.

Use Case

- sequence diagram capturing roles involved in the use case

the system process is to provide a description of the interactions between the multiple entities from a high level for the system

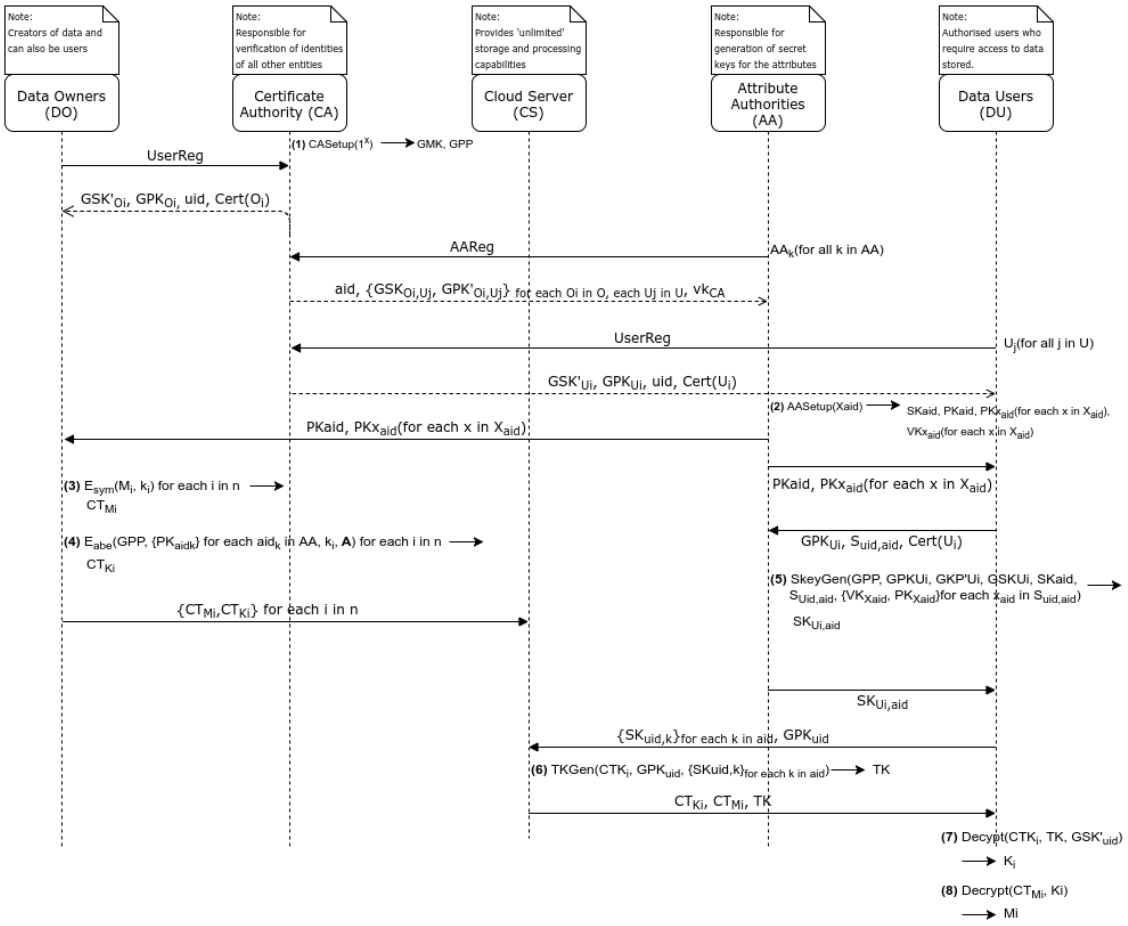

Figure 3.2: System Sequence Diagram

The sequence diagram is displayed in figure 3.2 showing a full process from system setup to encryption and decryption.

Chapter 4

Evaluation and Results

This chapter highlights the information about the software and hardware specifications involved in the implementation of the underlying framework and also shows the results of performance tests for the underlying ABE scheme in relation to scalability of computational and storage overhead with regards to the number of attribute authorities and attributes involved in the system for data encryption and decryption.

4.1 Experimental Setup

We evaluated the performance of our system framework by implementing the underlying ABE scheme in an environment using hardware and software tools that will be described below.

Hardware

- Linux Server - We used a – Server running Ubuntu 16.04 LTS with – RAM size and – CPU speed.

Software Tools

- Python - We wrote our underlying code using Python 3.x as this gave us access to some other effective libraries that played a role in the robustness of our code and evaluation during our experiments.cite python + numpy + matplotlib

- Charm Crypto Library cite charm-crypto paper - We used the charm-crypto library for the implementation of the ABE components of our framework. Charm is a library implemented in python that has multiple contibutors in the cryptographic field and as a result gives access to easy to use functionalities related to pairing based public key encryption schemes such as ABE. Our scheme was implemented using a group size of – and an elliptic curve of type –.

System performance was evaluated through the implementation of the underlying scheme on a linux server and also with the use of different software tools be described below.

Give explanation of the specifications of my implementation i.e system hardware specs, language and library used, specification of elliptic curve that group is based on, etc

- Software

  - charm crypto library provide an overview of the charm library...look up the

paper published for the library for corresponding details for this section.....is info on setting up the library necessary? can be added as an index section I guess

- Details about Group size and other parameters eg code use pairing group SS512 look up this detail and add info eg curve type and size(elliptic curve),

- Hardware - the framework was implemented on a desktop CPU running OS with specifications

  - system hardware specs (RAM size, processor type and speed, CPU name)

4.2 Performance Evaluation

We evaluated the performance of our system by evaluating the two major operations - encryption and decryption - and how their performance scales in relation to the number of attribute authorities and attributes involved in the operations.

Our results for encryption as indicated in figures encryption show that our encryption system is identical in performance with the encryption algorithm of the adapted scheme for this framework [cite original paper].

Our results for decryption show a constant decryption time irrespective of the number of attribute authorities or attributes in comparison with the increase in decryption time of the adapted scheme [cite original paper]. This can be seen in charles and co.

These results indicate that the amount of time for decryption involving a single attribute will be the same as that involving a thousand or more attributes, keeping the computational overhead constant and allowing for the use of low computational

devices for the decryption process.

- add charts here showing the performance (i.e how long it took to run)

A. how encryption and decryption times scale in relation to the number of attribute authorities (use constant number of attributes in this case) this will be two charts)

B. how the encryption and decryption times scale in relation to the number of attributes involved in encryption and decryption (use constant number of attribute authorities in this case i.e the max number of attribute authorities) this will be two charts)

C. how revocation scales (i.e key update for users)?

How performance was evaluated i.e scalability + speed

- Computation costs

- Communication costs

4.3 Security Analysis

- correctness?

- look up other high level forms of validating scheme and also look at the security proof of the base scheme to reference also.

Chapter 5

Conclusion and Future Work

Conclusion and future work here

Talk about what my scheme has been able to achieve and where it could go in terms of efficiency and accountability

Bibliography

[1] A. Cavoukian, "A Guide to the Personal Health Information Protection Act," *Access*, no. December, 2004.

[2] L. Ibraimi, M. Asim, and M. Petković, "Secure management of personal health records by applying attribute-based encryption," *Proceedings of the 6th International Workshop on Wearable, Micro, and Nano Technologies for Personalized Health: "Facing Future Healthcare Needs", pHealth 2009*, pp. 71–74, 2010.

[3] S. Narayan, M. Gagné, and R. Safavi-Naini, "Privacy preserving EHR system using attribute-based infrastructure," *Proceedings of the 2010 ACM workshop on Cloud computing security workshop - CCSW '10*, p. 47, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1866835.1866845

[4] J. A. Akinyele, C. U. Lehmann, M. D. Green, M. W. Pagano, Z. N. J. Peterson, and A. D. Rubin, "Self-Protecting Electronic Medical Records Using Attribute-Based Encryption," *ePrint IACR org*, vol. 1, pp. 1–20, 2010. [Online]. Available: http://eprint.iacr.org/2010/565

[5] M. Barua, X. Liang, R. Lu, and X. Shen, "ESPAC: Enabling Security and Patient-centric Access Control for eHealth in cloud computing," *International*

Journal of Security and Networks, vol. 6, no. 2/3, p. 67, 2011.

[6] S. Alshehri, S. Radziszowski, and R. K. Raj, "Designing a Secure Cloud-Based EHR System using Ciphertext-Policy Attribute-Based Encryption."

[7] T. Hupperich, H. Löhr, A.-R. Sadeghi, and M. Winandy, "Flexible patient-controlled security for electronic health records," *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, p. 727, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2110363.2110448$\delimiter"026E30F$nhttp://dl.acm.org/citation.cfm?doid=2110363.2110448

[8] G. Hsieh and R.-J. Chen, "Design for a secure interoperable cloud-based Personal Health Record service," *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*, pp. 472–479, 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6427582

[9] M. Li, S. Yu, Y. Zheng, and S. Member, "Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption," *Tpds*, vol. 24, no. 1, pp. 131–143, 2013.

[10] A. Sahai and B. Waters, "Fuzzy Identity-Based Encryption," pp. 457–473, 2005.

[11] R. Kissel, "Glossary of Key Information Security Terms Glossary of Key Information Security Terms," *Nist*, vol. NISTIR 729, no. Revision 2, 2013.

[12] S. Pearson, *Privacy and Security for Cloud Computing*, 2013. [Online]. Available: http://link.springer.com/10.1007/978-1-4471-4189-1

[13] A. Cavoukian, "Privacy Impact Assessment Guidelines for the Ontario Personal Health Information Protection Act," no. October, 2005.

[14] "Understanding EHRs, EMRs and PHRsNo Title." [Online]. Available: https://www.infoway-inforoute.ca/en/what-we-do/digital-health-and-you/understanding-ehrs-emrs-and-phrs

[15] P. Mell and T. Grance, "The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology," *Nist Special Publication*, vol. 145, p. 7, 2011. [Online]. Available: http://www.mendeley.com/research/the-nist-definition-about-cloud-computing/

[16] A. Menezes, P. van Oorschot, and S. Vanstone, "Chapter 01: Overview of Cryptography," *Handbook of Applied Cryptography*, pp. 1–48, 1996.

[17] B. Lynn, "On The Implementation of Pairing-Based Cryptosystems," Ph.D. dissertation, 2007.

[18] A. Beimel, "Secure Schemes for Secret Sharing and Key Distribution," Doctor of Science, Israel Institute of Technology, 1996.

[19] A. Shamir, "How To Share a Secret," *Communications of the ACM (CACM)*, vol. 22, no. 11, pp. 612–613, 1979. [Online]. Available: http://doi.acm.org/10.1145/359168.359176

[20] G. Blakley, "Safeguarding cryptographic keys," *Afips*, p. 313, 1979. [Online]. Available: http://www.computer.org/portal/web/csdl/doi/10.1109/AFIPS.1979.98

[21] J. Benaloh and J. Leichter, "Generalized Secret Sharing and Monotone Functions," *Advances in Cryptology — CRYPTO' 88*, vol. 403, pp. 27–35, 1988. [Online]. Available: http://dx.doi.org/10.1007/0-387-34799-2{_}3

[22] M. Ito, A. Saito, and T. Nishizeki, "Secret sharing scheme realizing general access structure," pp. 56–64, 1989. [Online]. Available: http://dx.doi.org/10.1002/ecjc.4430720906

[23] B. Waters, "Ciphertext-Policy Attribute-Based Encryption : An Expressive , Efficient , and Provably Secure Realization," vol. 02, no. subaward 641, pp. 53–70, 2011.

[24] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," Cryptology ePrint Archive, Report 2010/351, 2010, http://eprint.iacr.org/.

[25] A. Shamir, "Identity-Based Cryptosystems and Signature Schemes," *Advances in Cryptology*, vol. 196, pp. 47–53, 1985.

[26] D. Boneh and M. Franklin, "Identity-Based Encryption from the Weil Pairing," *SIAM Journal on Computing*, vol. 32, no. 3, pp. 586–615, 2003.

[27] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based Encryption for Fine-grained Access Control of Encrypted Data," *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 89–98, 2006. [Online]. Available: http://doi.acm.org/10.1145/1180405.1180418$\delimiter"026E30F$nhttp://portal.acm.org/citation.cfm?doid=1180405.1180418

[28] J. Bethencourt and B. Waters, "Ciphertext-Policy Attribute-Based Encryption," 2007.

[29] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure,scalable ,and fine-grained data access control in cloud computing.pdf," *Ieee Infocom*, pp. 1–9, 2010.

[30] Y. Shucheng, C. Wang, K. Ren, and W. Lou, "Attribute based data sharing with attribute revocation," *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security - ASIACCS '10*, p. 261, 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1755688.1755720

[31] X. Liang, R. Lu, X. Lin, and X. S. Shen, "Ciphertext Policy Attribute Based Encryption with Efficient Revocation," pp. 1–21, 2011. [Online]. Available: http://bbcr.uwaterloo.ca/{~}x27liang/papers/abewithrevocation.pdf

[32] Y. Cheng, Z.-y. Wang, J. Ma, J.-j. Wu, S.-z. Mei, and J.-c. Ren, "Efficient revocation in ciphertext-policy attribute-based encryption based cryptographic cloud storage," *Journal of Zhejiang University SCIENCE C*, vol. 14, no. 2, pp. 85–97, 2013. [Online]. Available: http://link.springer.com/10.1631/jzus.C1200240

[33] M. Chase, "Multi-authority Attribute Based Encryption," *Proceedings of the 4th Conference on Theory of Cryptography*, vol. 4392, pp. 515–534, 2007. [Online]. Available: http://www.springerlink.com/index/10.1007/978-3-540-70936-7

[34] S. S. M. Chow, "New Privacy-Preserving Architectures for Identity- / Attribute-based Encryption," no. September, p. 129, 2010.

[35] M. Chase and S. S. M. Chow, "Improving privacy and security in multi-authority attribute-based encryption," *ACM Conference on Computer and Communications Security*, pp. 121–130, 2009. [Online]. Available: papers3://publication/uuid/7B75720C-27AE-4FE4-AB73-BBA0FC0BAA87

[36] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial In-*

telligence and Lecture Notes in Bioinformatics), vol. 6632 LNCS, pp. 568–588, 2011.

[37] K. Yang, X. Jia, K. Ren, and B. Zhang, "DAC-MACS: Effective data access control for multi-authority cloud storage systems," *2013 Proceedings IEEE INFOCOM*, pp. 2895–2903, 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6567100

[38] K. Yang and X. Jia, "Expressive, efficient, and revocable data access control for multi-authority cloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1735–1744, 2014.

[39] M. Green, S. Hohenberger, and B. Waters, "Outsourcing the Decryption of ABE Ciphertexts," *Proceedings of the 20th USENIX conference on Security*, pp. 34–34, 2011.

Appendix A

Appendix B

Appendix C

Appendix D