**Chapter 1**

**INTRODUCTION**

**1.1 Data sharing**

Data sharing becomes an exceptionally attractive service supplied by cloud computing platforms because of its convenience and economy. Data sharing the ability to share the same data resource with multiple applications or users, It implies that the data are stored in one or more servers in the network and that there is some software locking mechanism that prevents the same set of data from being changed by two people at the same time. Data sharing may also indicate the sharing of personal information on a social media platform.

There are two mechanisms for data sharing:

1. Substance information exchange forums (SIEFs), used for phase-in (existing) substances that have been pre-registered.

2. Inquiry, used for non-phase-in (new) substances and for phase-in substances that have not been pre-registered

Data sharing poses specific challenges in participatory monitoring initiatives, for example where forest communities collect data on local social and environmental conditions.

**1.2 Cloud computing**

Cloud computing is shared pools of configurable computer system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and helps the users focus on their core business instead of being impeded by IT obstacles.A **cloud service** is any service made available to users on demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's own on-premises servers.Based on the deployment model cloud services can be classified into 3 types:

1. Public cloud
2. Private cloud
3. Hybrid cloud

The **public cloud** is defined as computing services offered by third-party providers over the public Internet, making them available to anyone who wants to use or purchase them. They may be free of charge or sold on demand, allowing customers to only pay per usage for the CPU cycles, storage or bandwidth they consume. These are assumed to be untrusted services.

Eg: Amazon Elastic Compute Cloud (EC2), IBM's Blue Cloud, Sun Cloud,Google AppEngine and Windows Azure Services Platform

**Private cloud** refers to a model of cloud computing where IT services are provisioned over private IT infrastructure for the dedicated use of a single organization. A private cloud is usually managed via internal resources. The terms private cloud and virtual private cloud (VPC) are often used interchangeably. These are assumed to be semi-trusted or fully trusted services.

Eg: Microsoft Azure is an example of a public cloud.

**Hybrid cloud** is a cloud computing environment that uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between the two platforms.

Eg: Google Cloud Platform, Amazon web services .

Information privacy, or **data privacy** (or data protection), is the relationship between the collection and dissemination of data, technology, the public expectation of privacy, legal and political issues surrounding them.

**1.3 Data security**

Data securityhas consistently been a major issue in IT. Data security becomes particularly serious in the cloud computing environment, because data are scattered in different machines and storage devices including servers, PCs, and various mobile devices such as wireless sensor networks and smart phones. Data security in the cloud computing is more complicated than data security in the traditional information systems.

Privacy concerns exist wherever personally identifiable information or other sensitive information is collected, stored, used, and finally destroyed or deleted – in digital form or otherwise. Improper or non-existent disclosure control can be the root cause for privacy issues. A private cloud is a cloud in which only the authorized users can access the services from the provider.

**1.4 Trust on cloud**

In the **public cloud**, applications and data run on the same shared public pool of resources – available to anyone with the swipe of a credit card.

A **private cloud** is a cloud computing hardware and software platform that is dedicated to your organization.

Files stored in reliable cloud services are some of the most secure files you can have, provided you have good passwords. Google, Microsoft, and Amazon all provide reliable cloud services for consumer file storage.

What makes to trust them?

**Redundancy**. The service typically stores at least three copies of each piece of data, all in different places.

**Security**. Provided you take care of your credentials (e.g. your password), only you can access those files.

**Safe sharing**. If you want to give other people read access (or even allow them to write it), you can. No need to make a copy of the file, no need to mail them a thumb drive

**1.5 Access control**

Acess control is the selective restriction of access to a resource. In computer security, general access control includes authentication, authorization, and audit. A more narrow definition of access control would cover only access approval, whereby the system makes a decision to grant or reject an access request from an already authenticated subject, based on what the subject is authorized to access.

Authentication and access control are often combined into a single operation, so that access is approved based on successful authentication, or based on an anonymous access token.

Authentication methods and tokens include passwords , biometric scans, physical keys, electronic keys and devices, hidden paths, social barriers, and monitoring by humans and automated systems.

**1.6 Data deduplication**

**Data deduplication** is a specialized data compression technique for eliminating duplicate copies of repeating data. It is also known as intelligent data compression and single instance data storage.

Data deduplication technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent.

In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis.

As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk.

Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

Hence when the users want to outsource their sensitive data they have to implement access control in order to prevent unauthorised access to the data and to preserve data deduplication.

**Chapter 2**

**TECHNIQUE**

The techniques that are used in this project are:

1. Attribute-based encryption
2. Online/ofﬂine cryptography
3. Outsourcing computation
   1. **Attribute-based encryption**

Attribute-based encryption is a type of [public-key encryption](https://en.wikipedia.org/wiki/Public-key_encryption) in which the [secret key](https://en.wikipedia.org/wiki/Secret_key) of a user and the ciphertext are dependent upon attributes (e.g. the country in which he lives, or the kind of subscription he has). In such a system, the decryption of a ciphertext is possible only if the set of attributes of the user key matches the attributes of the ciphertext.

A crucial security aspect of attribute-based encryption is collusion-resistance: An adversary that holds multiple keys should only be able to access data if at least one individual key grants access.

Attribute-based encryption (ABE) is one of useful crypto- graphic primitives to realize ﬁne-grained access control, which has been widely adopted in cloud computing. In ABE, each user obtains a private key related to his attribute set or access policy. More speciﬁcally,two kinds of ABE have been deﬁned for access control system, that is,

1. key-policy ABE and
2. ciphertext-policy ABE.

In key-policy ABE, the policy for users are bounded in the private keys during the key issuing phase.

In ciphertext- policyABE,such policy is inserted and bounded in the ciphertext instead. Both kinds of ABE have found important application scenarios.

**USAGE**

Attribute-based encryption (ABE) can be used for log encryption.Instead of encrypting each part of a log with the keys of all recipients, it is possible to encrypt the log only with attributes which match recipients' attributes. This primitive can also be used for [broadcast encryption](https://en.wikipedia.org/wiki/Broadcast_encryption) in order to decrease the number of keys used.

**2.2 Online/offline cryptography**

A paradigm called hash-sign-switch based on Chameleon hashing functions for designing efficient online/offline signature schemes. An online/offline signature scheme consists of two phases and it can efficiently enable handover authentication in wireless networks. Before the message to be signed is known, the ﬁrst offline phase is performed. The second online phase is performed once the message is known, and it is supposed to be very fast. In the online/ofﬂine signature schemes based on the hash-sign-switch paradigm, one security ﬂaw is the key exposure problem of Chameleon hashing.To solve this problem, a special double-trapdoor hash family based on the discrete logarithm assumption,and they applied the hash-sign-switch paradigm to propose a much more efﬁcient generic online/ofﬂine signature scheme.

In the proposed scheme, the encryption process is split into two phases: the ofﬂine phase and the online phase.The ofﬂine phase does the vast majority of the work to encrypt a message, and it requires neither the knowledge of the message to be encrypted nor the receiver’s identity. This division of computational tasks makes encryption affordable by mobile devices with limited computation power in that most of the works can be executed ofﬂine.

A new notion called identity-based online/ofﬂine key encapsulation mechanism, which allows the key encapsulation process to be split into ofﬂine and online stages. The ﬁrst fully secure online/ofﬂine predicate encryption and attribute-based encryption schemes have recently been presented by Datta et al. (2015). These schemes are chosen-plaintext attacks (CPA)secure and the computation efﬁciency on the data owner’s side still needs to be improved. Besides, these schemes fail to realize public ciphertext test before performing expensive decryption.

**2.3 Outsourcing computation**

With the rapid development of cloud computing, the enterprises and individuals can outsource their sensitive data into the cloud server where they can enjoy high quality data storage and computing services in a ubiquitous manner. This is known as the outsourcing computation paradigm.

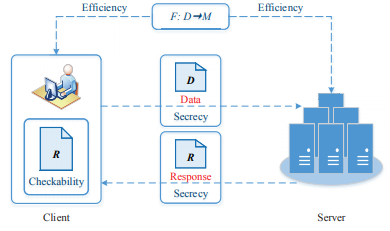


Figure 2.3.1: Secure outsourcing paradigm in cloud computing.

The issue of secure outsourcing computation, including the scientiﬁc computation problem and other computational problem, has drawn much attention for decades. Especially, with the fast development of mobile device and other resource-limited computational devices, such an outsourcing technique becomes more and more important. However, the previous techniques cannot be applied to ABE directly because of the problems are totally different. Some works also proposed to accelerate the computation of exponentiations. Recently,the technique of fully homomorphic encryption has also been utilized to construct and solve the problem of outsourcing computation.

**Chapter 3**

**ADVANTAGES OF ABE**

The criteria of an ideal Attribute-based Encryption:

1. **Data confidentiality**: Data Confidentiality is a set of rules or a promise that limits access or places restrictions on certain types of information. In cloud, the data was encrypted by the data owner and unauthorized parties including the cloud cannot know the information about the encrypted data hence data confidentiality is maintained.

2**. Secured access control**: Secured Access Control is any mechanism by which a cloud system grants or revokes the right to access some data, or perform some action. In cloud users are granted with different access right to access data to provide security.

3**. Scalability**: Scalability is defined as the capability to handle the user load supported, the number of transactions, the data volume etc. When the authorized users increase, the system can work efficiently. So the number of authorized users cannot affect the performance of the system.

4. **User accountability**: If the authorized user is dishonest, he would share his attribute private key with the other unauthorized user. It causes the problem that the illegal key would share among unauthorized users.

5**. User revocation**: If the user quits the system, the scheme can revoke his access right from the system directly. The revocable user cannot access any stored data, because his access right was revoked.

6. **Collision resistant**: Users cannot combine their attributes to decipher the encrypted data. Since each attribute is related to the polynomial or the random number, different users cannot conclude each other.

**Chapter 4**

**PROBLEMS AND CHALLENGE**

**4.1 Problems**

* The problem of achieving fine-grainedness still remains unresolved, granularity is the extent to which a material or system is composed of distinguishable pieces or grains.
* It can either refer to the extent to which a larger entity is subdivided, or the extent to which groups of smaller indistinguishable entities have joined together to become larger distinguishable entities.
* For example, a kilometer broken into centimeters has finer granularity than a kilometer broken into meters.
* Coarse-grained materials or systems have fewer, larger discrete components than fine-grained materials or systems.
* A coarse-grained description of a system regards large subcomponents while a fine-grained description regards smaller components of which the larger ones are composed.
* The concepts granularity, coarseness, and fineness are relative, used when comparing systems or descriptions of systems, The terms fine and coarse are used consistently across fields, but the term granularity itself is not.
* The problem of achieving scalability also still remains unresolved, Scalability is the capability of a system, network, or process to handle a growing amount of work, or its potential to be enlarged to accommodate that growth.
* The problem of simultaneously achieving high efficiency on the data owner’s side still remains unresolved, Efficiency is the (often measurable) ability to avoid wasting materials, energy, efforts, money, and time in doing something or in producing a desired result. In a more general sense, it is the ability to do things well, successfully, and without waste.
* Efficiency is very often confused with effectiveness. In general, efficiency is a measurable concept, quantitatively determined by the ratio of useful output to total input, Efficiency can often be expressed as a percentage of the result that could ideally be expected.
* The problem of achieving standard data confidentiality of cloud data sharing actually still remains unsolved, Confidentiality involves a set of rules or a promise usually executed through confidentiality agreements that limits access or places restrictions on certain types of information.
* Confidentiality has been defined by the International Organization for Standardization (ISO) in ISO-17799 as "ensuring that information is accessible only to those authorized to have access" and is one of the cornerstones of information security. Confidentiality is one of the design goals for many cryptosystems, made possible in practice by the techniques of modern cryptography.
* Confidentiality of information, enforced in an adaptation of the military's classic "need-to-know" principle, forms the cornerstone of information security in today's corporations. The so called 'confidentiality bubble' restricts information flows, with both positive and negative consequences.

**4.2 Challenges**

* If we take today’s scenario, the use of technology and internet made us relax in portability of data. We can nowadays share almost everything others like pictures, movies, thoughts, etc.
* If we need an emergency help from a doctor or hospital for chronic diseases like cardio, hepatic, neuro related previous data, we are now able to produce with the help of internet or cloud technology.
* Due to large number of internet users, it is also required to protect our data from being misused. An unauthorized person should not be made access to the private data of an individual.
* For this reason we are required to take care of data by implementing data protection techniques like cryptography. Our work in this area is based on Attribute based Encryption.

Key generation

Manage Manage

Data owner

user

Encrypted data User Access

Data storage

Fig 4.2.1: Architecture of data

* We are using the cipher text policy attribute based encryption that provides us to define set of attributes required by the decryptor for the cipher text.

* Everyone using the technique based on policy decrypts the required data. By this method, we didn’t need data server to prevent data from unauthorized user access.\
* There are many challenges, in using the attribute based encryption system in data sharing environment.
* The key generation process gives an approach for reduction of processing and authorization certification.
* We are proposing a model to work on attribute based encryption technique for the shared data. Proposed model works as shown in the fig. 1.
* Here the data storage contains the shared data which can be accessed by data owner and user.
* The data is managed by the key generation module which generates a key for encryption.
* The key generation module manages the data owner and user domains.
* The policy is defined in the key generation module which is implemented to manage the two modules viz data owner and user

Other main challenges are:

* Key coordination
* Key escrow
* Key revocation

**Chapter 5**

**PROPOSE MODEL AND HELP FUNCTION**

**5.1 Propose model**

The system architecture of attribute- based data sharing suitable for resource-constrained users in cloud computing consists of four entities AA (**Attribute Authorit**y), CSP (**Cloud Service** **Provider**), MO (**Mobile Data Owner**), and DU (**Data User**).

• AA is a key entity who generates system public parameters and master keys. Especially, the system public parameters contain immediate cipher texts, which can be used by MO in the online phase. Also, AA manages users in the system and it is fully trusted by entities in the attribute-based data sharing system.

• MO is a resource-constrained entity who wishes to safely store a ﬁle on cloud storage servers maintained by CSP for sharing. Before it speciﬁes the message, MO can generate offline cipher texts while accessing the power source. When the message becomes known, MO can calculate ﬁnal cipher texts online without signiﬁcantly draining the battery. • CSP is in charge of saving the cipher text data of MO and it consists of a lot of cloud storage servers, which are main- tained by a data service manager.

• DU is an entity who has a secret key and intends to access a cipher text hosted in CSP. In order to improve the efficiency of decryption, a public cipher text test phase is additionally introduced before the decryption phase. To be speciﬁc, after downloading the cipher text from CSP, DU should perform the test that if the cipher text is legitimate. And, the decryption phase is performed if and only if the cipher text passes the test.

•The security goal is semantic security of data, which is closely related to the following two requirements:

• **Data Conﬁdentiality**: Unauthorized users should be prevented from recovering the message of ciphertexts. In addition, unauthorized access from CSP to the message of ciphertexts should also be prevented.

• **Collusion-Resistance**: Malicious users colluding with CSP should not succeed in decrypting the ciphertext by combining their attributes if each of them cannot decrypt the ciphertext alone.

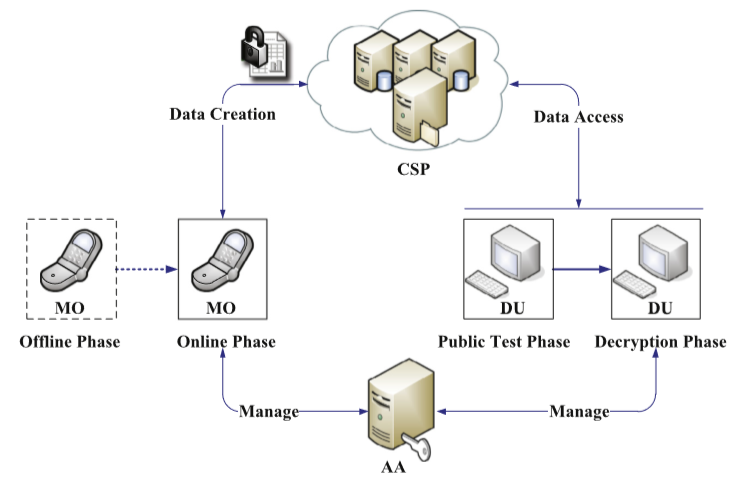


Figure 5.1.1: System architecture of attribute-based data sharing for resource-constrained users

Based on the proposed system architecture, we deﬁne the attribute-based data sharing system suitable for resource- constrained users in cloud computing. The system involves ﬁve phases as below.

•**Initialization**. AA generates system public parameters and master keys for the system. All users can obtain the system public parameters, where immediate cipher texts are calculated by AA and used in the subsequent online data creation phase by MO.

•**User Registration**. A user can join the attribute-based data sharing system by committing an access structure to AA, who issues a secret key to the user based on the access structure.

•**Ofﬂine Data Creation**. MO generates ofﬂine cipher texts, which are used in the subsequent online data creation phase by MO.

•**Online Data Creation**. MO encrypts a ﬁle based on an attribute set and outsources the ﬁnal cipher text to CSP for sharing.

•**Data Access**. DU downloads a cipher text from CSP. If the cipher text is legitimate, then MU decrypts it based on his/ her secret keys.

**5.2 Security model**

Before giving the formalized security model, we first lay out the definition of an online / offline KP-ABE scheme supporting public cipher text test, which is the basic tool of the proposed attribute-based data sharing system. An online/offline KPABE scheme supporting public cipher text test consists of the following five algorithms:

* **Setup**(1*ʎ*)→(*PK*,*MK*):The setup algorithm is run by AA. It takes as input a security parameter *ʎ*, and outputs the system public key *PK* and the master key *MK*.
* **KeyGen**(*PK MK*, , A) → *SK*A : The key generation algorithm is run by AA. It takes as input the system public key *PK*, the master key *MK* and an access structure A, and outputs *SK*A as the secret key associated with A.
* **Encryptoff** (PK) → *CT*off : The offline encryption algorithm is run by MO. On input the system public key *PK*, it generates an offline ciphertext *CT*off .
* **Encrypton**(*PK, m, S, CT*off ) → *CTS* : The online encryption algorithm is run by MO. On input the system public key *PK*, a message *m*, an attribute set *S* and an offline ciphertext *CT*off , it generates the final ciphertext *CTS*.
* **Decrypt**(*PK,CTS ,SK*A) → *m* or ⊥: The decryption algorithm is run by DU. It involves a public ciphertext test phase and a decryption phase. Note that the secret key *SK*A is only used in the decryption phase. On input the system public key *PK*, a ciphertext *CTS* of a message *m* under *S*, and a secret key *SK*A associated with A, the ciphertext *CTS* is tested and decrypted by DU as follows:

1. *Public Test Phase*: It returns ⊥to terminate decryption if *CTS* is illegitimate. Otherwise, the *Public Test Phase* ends by initiating the *Decryption Phase*.
2. *Decryption Phase*: It outputs the message *m* if *S* is an authorized set of A.

**5.3 Public Cipher Text:**

In public cipher test anyone can verify whether a cipher text is legitimate without requiring secret keys. Invalid cipher texts are thrown away without performing the decryption phase. A public cipher text test phase is performed before the decryption phase, which eliminates most of the computational cost resulted from illegitimate cipher texts. In other words, the public cipher text test allows a user to check at a low cost whether a potential equation holds for components of a given cipher text before performing the expensive decryption phase.

**5.4 Help Function**

**5.4.1 Chameleon hash function**

A chameleon hash function is a trapdoor collision-resistant hash function, which is associated with a trapdoor/hash key pair. Anyone who knows the public key pkch can efﬁcientlycompute the hash value for each input.

The technique of Chameleon hash function is used to generate an immediate cipher text, which will be blinded by the ofﬂine cipher texts to obtain the ﬁnal online cipher texts. In this way, the proposed scheme is proven CCA2 secure,which is widely recognized as a standard security notion. Theoretical analysis and experimental results indicate that the proposed ABDS system is extremely suitable for resource imited mobile users in cloud computing.

**Chapter 6**

**Conclusion**

Aiming at tackling the computation efficiency and weak data security issues in cloud data sharing, we propose an attribute based data sharing scheme suitable for resource-limited mobile users in cloud computing. The proposed scheme supports online/offline encryption modes and allows anyone to check the validity of ciphertexts before expensive full decryption. Even the computation task in offline phase is signiﬁcantly reduced by adding system public parameters.

The Project Phase-I report is hereby submitted and further the implementation part is to be carried out.