Hierarchical based Cloud Computing and De-Identification of Secured Big Data

C. Heartina Heaveny
M. Tech Student, Dept. of CSE
Meenakshi College of Engineering
Anna University Chennai, India

V. Sarala, PhD
Asst. Professor, Dept. of CSE
Meenakshi College of Engineering
Anna University Chennai, India

J.C. Kavitha
HOD, Dept. of CSE
Meenakshi College of Engineering
Chennai-600 078

ABSTRACT
A smart grid is modernized electrical equipment that uses analog or digital information in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity. The main challenges of smart grids, however, to manage different types of front-end intelligent devices such as power assets and smart meters efficiently, and how to process a huge amount of data received from these devices. Cloud computing is a model for enabling ubiquitous network access to a shared pool of configurable computing resources. It relies on sharing of resources to achieve coherence and economies of scale. This Cloud computing technology is used to address the challenges of Smart grid. The system consists of secure cloud computing based framework for big data information management in smart grids. It has a hierarchical structure of cloud computing center to provide different types of computing services for information management and big data analysis. In addition to this structural framework, a security solution is proposed based attribute based encryption to address critical security issues of the proposed framework.

Keywords
Smart Grid, Attribute Based Encryption, Cloud.

1. INTRODUCTION
In order to allow intelligent power control and monitoring, the concept of smart grid has been gaining tremendous attention among both researchers and utility providers recently. Specifically, in smart grid, advanced technologies (i.e., sensing, control, digital communication, and network information), as shown in table 1, are merged with power system engineering to effectively address numerous critical issues that limit existing electricity grids, such as the lack of adequate demand response, scalability, energy conservation, reduction of carbon emission, and control of distribution. Smart grid is perceived to transform the energy industry by allowing bidirectional communication between the consumers and the energy producers and/or operators. The necessary devices required for facilitating this two-way communication consist of smart meters [1] [3]. Smart meters, along with power instrumentation and monitoring sensors, form the core of advanced metering infrastructure (AMI) aimed at maintaining high levels of performance, reliability, and manageability. Nevertheless, researchers have expressed their deep concern about the need to integrate a security infrastructure with the smart grid's AMI [4] [5]. The reason behind this concern is the fact that many smart grid projects are widely using advanced communications networks to link numerous sensors and smart meters for connecting consumers with the utility providers so as to exchange information bi-directionally. However, as evident from the wide variety of malicious threats against existing communication networks such as the Internet, the smart grid communication framework is expected to increase the vulnerability of the grid to cyber-attacks [1, 2] such as denial of service attacks, spoofing, privacy leakage, and so forth. As a result, the development of a wide range of communication networks for supporting the integration of open access energy competition through AMI should set out adequate security provisions for protecting the smart grid communication.

Table 1: The Transformation of the Traditional One-way-Communication Power Grid to the Two-way-Communication Smart Grid

Traditional grid	Smart grid
Electric machinery	Digital
One-way communication	Two-way communication
Centralized power generation	Distributed power generation
A small number of sensors	Full grid sensor layout
Manual monitoring	Automatic monitoring
Manual recovery	Automatic recovery
Failures and power outages	Adaptive and Islanded
Few user options	More user options

1.1 Cloud Computing and Smart Grid
Cloud computing is of interest to the power community for several business reasons. Some parallel the green energy considerations that have stimulated such dramatic change in the power industry: cloud computing is a remarkably efficient and green way to achieve its capabilities. Others reflect pricing: cloud computing turns out to be quite inexpensive in dollar terms, relative to older models of computing. Also, cloud computing offers astonishing capacity and elasticity. It satisfies the requirement for Smart Grid security objectives. It provides the following,

- Scalable, Consistent and fault tolerant real time services
- Protection of Privacy data
- Decentralization
- Reliability and Agility
- Energy and cost efficiency

1.2 The Cloud Cost Advantage
The Smart Grid needs a national-scale, pervasive network that connects every electricity producer in the market, from coal and nuclear plants to hydroelectric, solar, and wind farms, and small independent producers, with every electricity consumer, from industrial manufacturing plants to residences, and to every device plugged into the wall. This network should

enable the interconnected devices to exchange status information and control power generation and consumption. The scale of such an undertaking is mind boggling. Yet, the key enabler, in the form of the network itself, already exists. Indeed, the Internet already allows household refrigerators to communicate with supermarkets and transact purchases. It won't be difficult to build applications ("apps") that inform the washing machine of the right time to run its load, based on power pricing information from the appropriate generators. Whatever their weaknesses, the public Internet and cloud offer such a strong cost advantage that the power community cannot realistically ignore them in favour of building a private, dedicated network for the smart grid.

2. LITERATURE SURVEY AND RELATED WORK

2.1 Secure Cloud Computing Framework for Big Data Management of Smart Grid

It consists of a hierarchical structure of cloud computing centre to provide different types of computing services for information management and big data analysis. The structure consist of top, regional and end user levels, where the top level is involved in managing the services across the regional cloud computing centres. The algorithm used for encrypting the Smart grid information/file is Identity Based Encryption and Identity Based proxy Re-encryption algorithms. The data is encrypted using the Identity of the end user and re-encrypted by the Information storage using the proxy re-encryption algorithm for storage. A private key generator is used for allocating the private key for each end user. The Top management level users view the data for performance analysis. [1]

2.2 Identity Based Cryptography for Grid Security

The system provides identity based key that secures the job submission for gird users. It provides Grid security infrastructure. The algorithm used is Hierarchical Identity Based Encryption (HIBE). In this algorithm, the private key size increases as it descends down the hierarchy tree (organization hierarchy). An identity level K of the hierarchy tree can issue private key to its descendants but cannot decrypt the messages intended from any other identities. It does not support public key infrastructure. It's comparatively lower cost and it supports single sign on mode. SSO mode is a concept, in which one single login credentials are used for any application. It is generally used for deployment or testing phase in the system by the admin. [2]

2.3 Running Smart Grid Control Software on Cloud Computing Architecture

The system provides a cloud environment for the Smart grid information available, with a centralized storage and access to the information when required. The system explains the benefits of running the Smart grid information in Cloud environment. It consists of number of benefits including the security, scalability, availability and reliability. It provides centralized storage in order to access information from any location. It supports Checkpoint barriers. Checkpoint barriers, are used generally, when one node has failed, the entire nodes in that particular regions fails. But with this checkpoint, only some of the nodes within the checkpoint restart. But implementing this checkpoint barrier in cloud is costly. [3]

2.4 An Identity based Security Infrastructure for Cloud Environment

In this system, a security infrastructure for service-oriented Cloud applications that overcomes the problems of previous Grid security solutions and allows complex applications to be deployed and used on-demand in a secure manner. It makes use of a novel identity-based cryptographic (IBe) system to avoid the complexity and management problems of certificate-based security infrastructures. With an IBE approach, any string can be used as a public key. This leads to a much more intuitive security environment without the need for a heavyweight certificate infrastructure. Not only is it unlikely that all Cloud resource providers will want to cooperate or have a single root of trust, but this drawback of IBE systems creates a significant management overhead that defeats the benefit of easy usage promised by IBE. This system provides features for signing and encrypting the messages that are exchanged between the web services and the clients. [5]

2.5 Secure Information Aggregation for Smart Grid using Holomorphic Encryption

The system supports data aggregation from all the smart meters in any given area. With aggregation tree the route covers the arbitrary set of designated nodes. The encryption algorithm represents a group of semantically secure encryption function that allow certain algebraic operations on plain text to be performed directly on cipher text. The system supports resistant to Dictionary attack (same data will be encrypted to different cipher with different blinding factor). It is also in deterministic in nature. Since all the information of different regions are stored in centralized location, it causes excessive traffic flow in networks. Also, the use of data aggregation causes to change in the data report generated by this aggregation. [6]

2.6 Identity Based Authentication Scheme in Cloud Computing

The identification presented in this system, is based on public key cryptography technologies. It presents non authentication centre scheme, which avoids the Key escrow and the Key revocation problem in the authentication scheme based on public key certificate and then the security scheme is analysed, firstly, the legitimacy of user identity should be confirmed, and then the cloud server and the user will negotiate the session key, then a secure data transmission channel is established. It is based on Diffie-Hellman problem and the security is proved. Key escrow is a backup of cryptographic key held by the third party, the system does not support the key escrow problem. The algorithm avoids Key revocation, in which the key is retired by the PGP permanently. The system proves the user's legacy and the system should know if the user has right to access. The collusion of users occurs in the system due to identity mismatch. [8]

2.7 Improved Proxy Re-Encryption with Applications to Secure Distributed Storage

In this system, cryptosystem has two stages decryption procedure with two different secret keys. It minimizes the users' secret storage and hence it is key optimal. Also the global system parameters remain unchanged. It also provides empirical performance measurements of applications using

proxy re-encryption. To demonstrate the practical utility of our proxy re-encryption schemes, we measure an implementation of proxy re-encryption used in a secure file system. The system uses a centralized *access control server* to manage access to encrypted content stored on distributed, untrusted replicas. It supports non interactivity (since the delegatee does not need to be involved in the generation of the re-encryption keys). It has proxy invisibility (The proxy in the BBS scheme is *transparent* in the sense that neither the sender of an encrypted message nor any of the delegatee have to be aware of the existence of the proxy) It also uses proxy re-encryption to allow for centrally-managed access control without granting full decryption rights to the access control server. [10]

2.8 Identity Based Encryption from Weil Pairing

In this system, a fully functional identity-based encryption scheme was proposed. The performance of this system is comparable to the performance of ElGamal encryption. The security of the system is based on a natural analogue of the computational Diffie-Hellman assumption. Based on this assumption, the new system has chosen cipher text security in the random oracle model. Using standard techniques from threshold cryptography the PKG, can be distributed so that the master-key is never available in a single location. The system supports non degeneration of public key. It does not provide support against any of the CCA attack. A single escrow key enables the decryption of cipher text encrypted under any public key.

3. EXISTING SYSTEM

The existing system consist of a hierarchical structure of cloud computing centre to provide different types of computing services for information management and big data analysis. The structure consist of top, regional and end user levels, where the top level is involved in managing the services across the regional cloud computing centres.

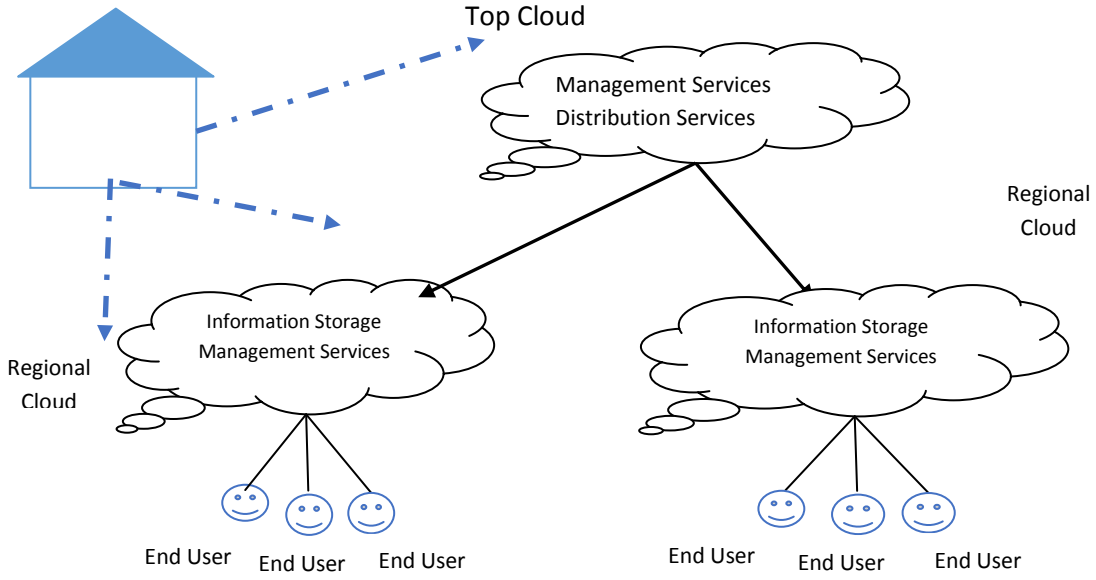


Fig 1: Illustrates Hierarchical Architecture

The regional cloud computing centres has the Information Storage that stores all the information from the End user level. Any information is transferred only from low level to higher level. In each level, every user and regional cloud service provider is provided with their own Identity which is used for encrypting the data sent by the end user and stored in information storage. When the information is requested by the Top level, the Information storage re-encrypts the data with the Top level identity as Proxy identity using the Identity Based-Proxy Re-encryption Algorithm.

Identity Based Proxy Re-encryption lets the third parties to change the cipher text which is encrypted by one party to decipher by another party. In the existing system, the proxy transforms the encryption under Alice's identity to be encrypted by Bob's Identity without gaining any information about the plain text. It is a combination of Identity based encryption and Proxy re-encryption. IBE schemes are more costly compared to other ElGamal style encryptions due to the pairing operations.

A PKG (Private Key Generator) issues private keys to Top level, regional level and end users when they are registered using their unique strings. These unique strings include, email id, address etc. These keys are used for encryption or signature verification keys. Each entity will send data to a higher level user. That is, the end users send information to the regional cloud and Top cloud only. Each entity authenticates the data using the private key generated by PKG.

4. PROPOSED SYSTEM

The proposed system will be an encryption algorithm that overcomes the drawbacks of Identity Based encryption (IBE) and Identity based Proxy Re-encryption algorithm. It uses the Attribute Based Encryption algorithm and CCA Secure Proxy Re-encryption algorithm.

The ABE algorithm involves, encrypting the data using the attributes defined by the developer. An access structure is defined for each entity in the framework and an attribute tree is developed. The leaf nodes consist of the attributes, the logical gates of access tree is the node. A secret key is generated using the access structure depending on the defined attributes.

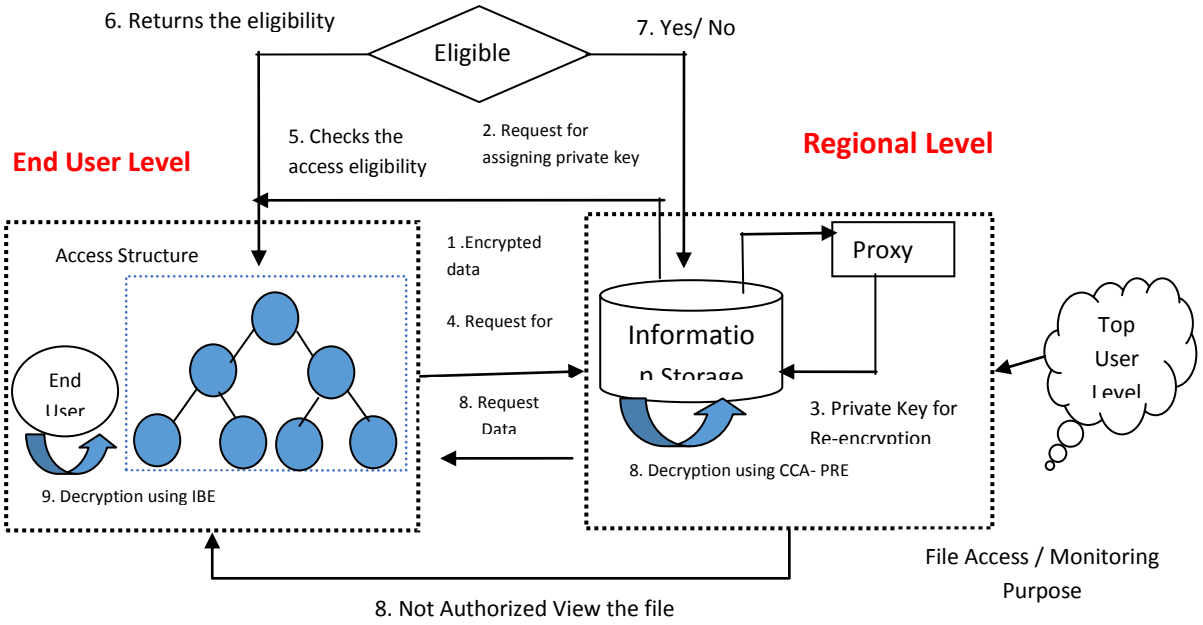


Fig 2: Proposed System Architecture

The de-identification can be defined as replacing, removing or hiding any data that requires more security. The reciprocal f this operation is named as re-identification of data. The process involves 5 main steps, Tokenization, Codification, Detection, Storage and Replacement.

5. CCA PROXY RE-ENCRYPTION

In a proxy re-encryption scheme, a semi-trusted proxy can transform a cipher text under A's public key into another cipher text that Bob can decrypt. However, the proxy cannot access the plaintext. Due to its transformation property, proxy re-encryption can be used in many applications, such as encrypted email forwarding. [11]

In this paper, by using signature of knowledge and Fijisaki-Okamoto conversion, we propose a proxy re-encryption scheme *without* pairings, in which the proxy can only transform the cipher text in one direction. The proposal is secure against chosen cipher text attack (CCA) and collusion attack in the random oracle model based on Decisional Diffie-Hellman (DDH) assumption over $Z^{*}{}_{N}{}^{2}$ and integer factorization assumption, respectively. Hence, the CCA-secure unidirectional PRE schemes without pairings are desired. [11]

6. DE-IDENTIFICATION OF BIG DATA

The de-identification can be defined as replacing, removing or hiding any data that requires more security. The reciprocal f this operation is named as re-identification of data. The process involves 5 main steps,

- *Tokenization* – this method separates the full data available and the data that needs to be secured
- *Codification* – this method assigns an ASCII value to the each letter in the data to be secured. Eg – [J A C K] = {70,110,30,23}
- *Detection* - Algorithm is as follows,

For each word in the sentence W do

For each letter in the word L do

Compute D the distance between W and L

If (D< threshold) then pass to storage step

Else pass to the next letter

End for

End for

- *Storage* – This method use the map-reduce concept. The changed word is stored in the memory and it gets updated if the value of word changes
- *Replacement* – This method replaces the original word from the ASCII value whenever the data is requested for.

7. ALGORITHMS

The KP-ABE scheme is carried out by four algorithms is described.

- *Setup Attributes*: This algorithm set attributes for each users. It does not take any input other than it contained security parameter for randomized algorithm. The bilinear group G1 with a generator g, with the prime order p a bilinear map

E.g.: $G1 \times G1 \rightarrow G2$.

The public key and master key determined for each user attributes.

They public key and master key are denoted by U, PK and MK.

Attributes- U = {1, 2. . . N}→ (1)

Public key- PK = (X, T1, T2 . . . , Ti) → (2)

Master key- MK = (x, t1, t2. . . , ti)→ (3)

Where Ti € G1 and ti € Zp are for attribute i, 1≤ i≤ N, and X € G2 is another public key component. We have Ti = gti and X = e (g, g) y, y∈ Zp. While PK is public key is used for

encryption, MK is master key kept as a secret by the authority party for secret key generation.

- *Encryption:* In this algorithm a message M, the public key PK, and a set of attribute U are taken as input. The is outputs the cipher text CT with the following format:

CT = (U, Ẽ, {Ei ∈ U}) → (4)

Where Ẽ= MKs, Ei = Tis. And s is randomly chosen from Zp

- *Key generation:* In this algorithm it takes access tree T as an input, with MK as the master key and PK as public key. Secret key SK is generated as outputs. First, it select a random polynomial pi(x) for each node i of T in the top-down manner starting from the root node r. For each non-root node j,

pj(0) = pparent(j)(idx(j))

where parent(j) represents j's parent and idx(j) is j's unique index given by its parent. For the root node r, pr(0) = y. Then it outputs SK as follows.

SK = {ski} i ∈ L (5)

Where L denotes the set of attributes attached to the leaf nodes of T and

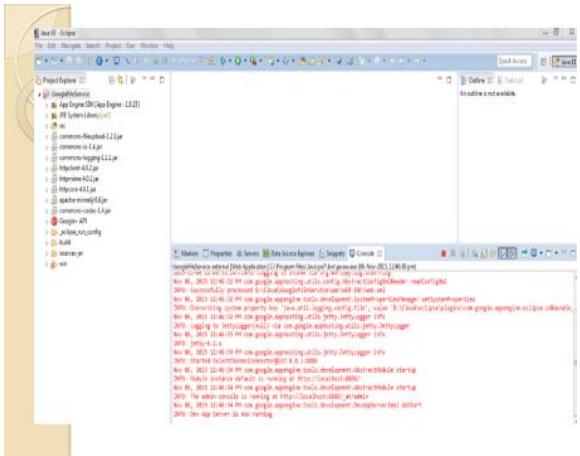
ski = gpi(0)/ti.

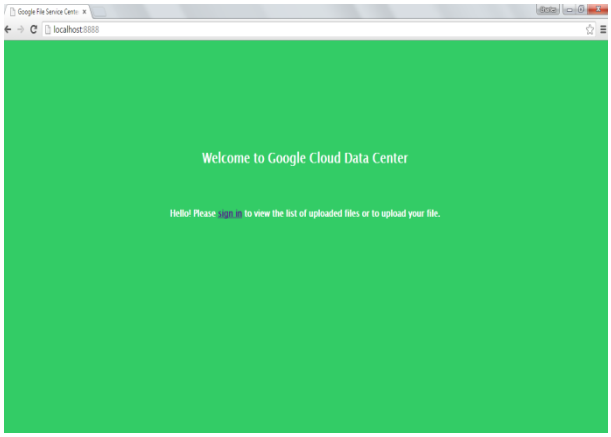
- *Decryption:* The encrypted cipher text E under the user's data attribute set I, the user's secret key SK with the access tree T with public key PK is takes as input for this algorithm. It first compare the

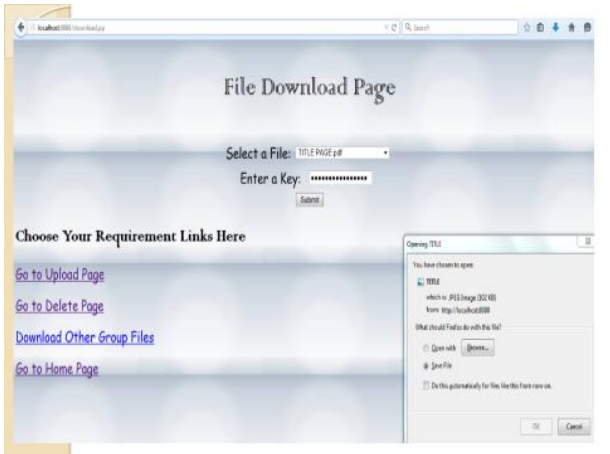
e(Ei, ski) = e(g,g)pi(0)s

For leaf nodes if it's satisfied then it aggregate the pairing results in the bottom-up by using polynomial interpolation technique. At last it recover the blind factor Ys=e(g,g)ys and decrypted message M if it's satisfied the data attribute I and access tree T.

8. EXPECTED OUTCOMES







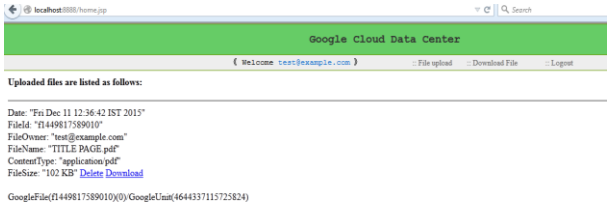


Fig. 3: Proposed Expected Outcomes

9. CONCLUSION

In this paper, the proposed Smart Grid information management system is put together with cloud environment to overcome the drawbacks in security of Smart Grid. The algorithm proposed in the first phase provides, secured transmission of data between the various cloud entities. The security provided with the algorithm is costly because of using PKI. Also, double encryption provides identity based encryption and corresponding proxy re-encryption. The repudiation of data, collusion of end users is overcome in the next phase of the project that implements the Attribute Based encryption and CCA [11] secure proxy re-encryption algorithm.

10. REFERENCES

[1] Joonsang Baek, Quang Hieu Vu, Joseph K. Liu, Xinyi Huang, and Yang Xiang, "A Secure Cloud Computing Based Framework for Big Data Information Management of Smart Grid". IEEE Transactions on Cloud Computing, vol. 3, no. 2, Apr/Jun 2015.

[2] H. Lim and K. G. Paterson, "Identity-based cryptography for grid security", Int. J. Inf. Security, vol. 10, no. 1, pp. 15–32, 2011.

[3] K. P. Birman, L. Ganesh, and R. V. Renesse, "Running smart grid control software on cloud computing architectures", in Proc. Workshop 2011, pp. 1–33.

[4] H. Khurana, M. Hadley, N. Lu, and D. Frincke, "Smart-grid security issues", IEEE Security Privacy, vol. 8, no. 1, Feb. 2010.

[5] C. Schridde, T. D. Ornemann, E. Juhnke, B. Freisleben, and M.Smith, "An identity-based security infrastructure for cloud environments", in Proc. IEEE Wireless Commun., Netw. Inf. Security, 2010, pp. 644–649.

[6] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption", in Proc. IEEE Conf. Smart Grid Commun., 2010, pp. 327–332.

[7] S. Rusitschka, K. Eger, and C. Gerdes, "Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain", in Proc. 1st Int. Conf. Smart Grid Commun., 2010.

[8] H. Li, Y. Dai, L. Tian, and H. Yang, "Identity-based authentication for cloud computing", in Proc. 1st Int. Conf. Cloud Comput., 2009, vol. 5931.

[9] M. Green and G. Ateniese, "Identity-based proxy re-encryption", in Proc. 5th Int. Conf. Appl. Cryptograph. Netw. Security, 2007.

[10] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage", ACM Trans. Inf. Syst. Security, vol. 9, no. 1, pp. 1–30, 2006.

[11] Jun Shao, Zhenfu Cao, "CCA-Secure Proxy Re-Encryption without Pairings", www.iacr.org/54430361.pdf