Towards security in sharing data on cloud-based social networks

4 authors, including:

Duc H. Tran
Nanyang Technological University

5 PUBLICATIONS   **18** CITATIONS

SEE PROFILE

Hai-Long Nguyen
Nanyang Technological University

6 PUBLICATIONS   **35** CITATIONS

SEE PROFILE

Wee Keong Ng
Nanyang Technological University

382 PUBLICATIONS   **3,828** CITATIONS

SEE PROFILE

Towards Security in Sharing Data on Cloud-Based Social Networks

Duc H. Tran, Hai-Long Nguyen, Wei Zha, Wee Keong Ng

School of Computer Engineering

Nanyang Technological University, Singapore 639798

Email: {ductran, longnguyen, zhawei, wkn}@pmail.ntu.edu.sg

Abstract—Nowadays, ubiquitous sensor devices such as mobile phones, laptops, GPSes, etc. allow one to access not only his own data but also others' on cloud servers. To enhance security, data is usually encrypted before sending to the servers. However, making use of others' ones encrypted data without decryption keys is very challenging. In this paper, we suggest a framework that allows users of cloud-based social networks to share their private data in a secure manner. In our framework, every user in a group has his own secret key to encrypt and decrypt data. The key will be revoked if the user leaves the group. Using proxy re-encryption schemes, the framework helps any user be able to access others' data in the same group.

I. INTRODUCTION

The demand of outsourcing data has greatly increased in the last decade. To satisfy the need of data storage, many Storage Service Providers have appeared. Most of them are provided in the form of cloud services such as Amazon Simple Storage Service [13], Google App Engine [14], Microsoft Azure [16] and Dropbox [10]. Users and companies are able to access and store their data on the cloud. Storage services are accessed by not only traditional desktop computers but also ubiquitous sensor devices such as mobile phones, laptops. For instance, one may take a picture with his Apple iPhone[TM] and uploads to Dropbox for storage. In the same time, he sends the link of that picture to share with his friends. His friend can access the picture either by an Apple iPad[TM] or a laptop. By this way, a social network is build with data stored on the cloud.

A recent report by TheInfoPro [19] shows that nearly 20% of Fortune 1000 organizations stores their data on the cloud [8]. Storing data on the cloud brings following benefits: reliability, availability, fault-tolerance, better performance and cost efficiency. Imaging a small company uses its own PC to provide storage service. Once errors such as software crashes or system faults happen with the computer, all the services are immediately not available. Unfortunately, if the hard disk is broken, all the important data may be gone. While storing data on the cloud, they do not need to worry about these issues. Storage service providers take care all of the troubles for users. Apart from business data, personal data is also in the trend of moving to cloud. People are more willing to store their own data on the cloud to benefit the advantage of ubiquitously accessing their data with their smart devices.

While benefiting advantages from cloud computing, the disadvantages can not be ignored. The biggest disadvantage of storing data on cloud is data privacy. It can not be guaranteed that the storage service provider will not use the data for other purposes. A company is not allow its data to be sold to its competitors. That could cause the company lose its markets, revenues, reputation or even collapse. On the other hand, disclosing one's private data without permission is against the law [12]. For some privacy reasons, one may only want his data to be accessed by particular people. For instance, most people do not want to share their home address to strangers.

Traditional access controls are mostly used for in-house services only. For instance, a storage service provided by a company can manage users, deciding who can access into which data. However, when the storage services are moved to cloud, the traditional security models are no longer valid. Even some cloud services claim user can manage the accessing and sharing of their own data. They can assure that ordinary users cannot see your data, but the service provider's employees.

A trivial solution to benefit the advantages of cloud service while preserving privacy is to encrypt data before sending data to cloud for storage and when need, downloading data to local computer to decrypt it. Thus, even if the competitors see the data on the cloud, they cannot disclose real contents unless they have the key to decrypt the data. However, this solution is not practical since the data stored on cloud is often huge, decrypting such a large data for every query is impossible. Moreover, the data needs to be re-encrypted once the private key is leaking. For instance, an employee leaves his current company and join its competitor. All the encrypted data are not safe to that competitor any more. One possible solution is changing the private and public keys and encrypt the raw data again. But re-encrypt all the data may not be trivial since encryption and decryption in public-key schemes are both very expensive.

In this paper, we introduce a secure framework to efficiently share data among multi-users. In our proposed frame work, data encryption is also required before sending to cloud. However, the encryption and decryption are based on proxy re-encryption schemes. All the users use the same public key to encrypt data but different private key to decrypt data. Specifically, encrypted data will be pre-decrypted according to user's private key before sending to user. Thus, when a user left the company but trying to access the encrypted data again, simply not to pre-decrypt the data can keep the data safe. Without pre-decryption phases, user's private key is useless.

The rest of this paper is organized as follows. Preliminary

work including cloud storage services, social networks and data encryption schemes is introduced in Section 2. Section 3 presents a secure framework for data sharing on cloud-based social networks. We conclude the paper and discuss some future research directions in Section 4.

II. PRELIMINARIES

A. Cloud Computing & Social Network

Cloud computing, the prolonged dream of computing as a utility, has recently become a highly disruptive technology. The cloud computing can be considered as the next evolutionary step, including many related legacy technologies such as semantic web, large-scale distributed systems, autonomic computing. Cloud computing aims to provide computation, data access, software, and storage services that are available from anywhere, anytime, on demand and do not require end-user knowledge of the underlying system. Cloud computing is currently seized much attention with billion-dollar investments from many big companies such as Amazon EC2/S3, Microsoft Azure, Google App Engine, Saleforce.

In this decade, online social network has developed quickly and becomes an everyday part of most people's lives. There are hundreds of million people in many social networking sites such as Facebook, Twitter, MySpace, using blogs, and posting on YouTube and Flickr. Social networks model the real life relationships by providing facilities for users to communicate, interact and share their data via the Internet.

The number of users is still increasing daily, for example, the number of Facebook's user increases from 500 millions to 600 millions within six months, from July 21, 2010 to January 5, 2011 [20]. And the vast amount of produced information in the social networks has sparked the interest of cloud computing services, which can handle massive and fast growing data sets in a scalable and reliable manner with affordable cost. Traditionally, cloud computing provides platforms to host social networks or to develop scalable applications. For instance, Curry et. al. proposed an enterprise approach to integrate social networking and cloud computing by simulating and visualizing Facebook applications hosted by cloud providers. Facebook and Amazon Web Services are cooperating to help developers build instantly scalable applications for Facebook users [1]. Furthermore, social networking technology can be added into the cloud platform to help professional collaboration work. For example, Scribe has released Scribe Online, a cloud-based integration tool that makes of use of social collaboration technology [17].

B. Proxy Re-encryption

Privacy preserving is one of the key concerns of cloud computing as well as social network. With cloud services, users can put their personal data to the cloud and easily access it everywhere via the Internet. Together with the ease of use, users also crucially expect the cloud providers to supply secure mechanisms to authenticate users, secure the data in order to minimize the risk of identity theft and fraud. However, most cloud providers cannot guarantee that their services are fully secured.

In a social network, privacy covers not only the protection of personal information, for example, users' profiles, contacts and postings, but also the data sharing among a group of trusted members. For example, a group of Facebook users want to share pictures among themselves, they must totally rely on the Facebook service to set appropriate permissions.

Applying cryptography can help provide a solution for the privacy preserving problem in cloud-based social networks. Users only need to encrypt their data before uploading it to the cloud services, and keep their private key in a safe place. Moreover, the proxy re-encryption, a newly-devised cryptographic primitive, can be used to share data among users without relying on the cloud services. The goal of the proxy re-encryption is to securely enable the re-encryption of ciphertexts from one secret key to another key so that other member can use his private key to encrypt the ciphertext securely without worry about illegal users who can see the data.

We first review a proxy re-encryption that will be used in our framework for privacy preserving on cloud-based computing. This encryption scheme is constructed upon the ElGamal encryption scheme.

1) ElGamal Encryption: ElGamal encryption is a public-key cryptography system, invented by T. ElGamal in 1985 [11]. Its security is based on the difficulty of finding discrete algorithms modulo a large prime. The ElGamal system consists of three main steps:

- Initialization: Given a prime p, a primitive root r of p, the initilization step chooses randomly a secret key x with $0 < x < p$, and computes $b \equiv r^x \bmod p$. The public key is (p, r, b), and the secret key is x.
- Encryption: Given a plain text m, the encryption step chooses a random k with $0 < k < p$, then calculates $\gamma \equiv r^k \bmod p$ and $\delta \equiv b^k.m \bmod p$.
  The ciphertext corresponding to the plain text P is the ordered pair $\mathsf{Enc}(m) = (\gamma, \delta) = (r^k, r^{xk}.m) = C$.
- Decryption: The decryption step decrypts the ciphertext C with the private key a as follows:

$$\begin{aligned} \mathsf{Dec}(C) &= (\gamma)^{-x}.\delta \bmod p = (r^{xk})^{-1}.b^k.m \bmod p \\ &= (b^k)^{-1}.b^k.m \bmod p = m \bmod p. \end{aligned}$$

, where $(\gamma)^{-1} \bmod p$ is the inverse modulo of $\gamma \bmod p$.

2) ElGamal-based Proxy Re-encryption: Figure 1 illustrates the proxy encryption scheme. There is a pair of private keys that is created for each user and the proxy. Let us assume that a user i desires to share data m to a group. He first encrypts his data and send the his ciphertext to the cloud. When a user j requires to get the data from user i. The proxy will convert the ciphertext from the private key of user i to the ciphertext from the private key of user j. Thus, the user j can use his private key to decrypt and then gets the sharing data.

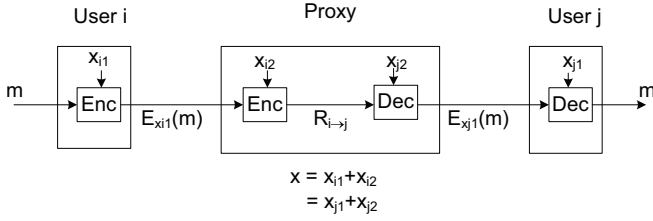In details, the ElGamal-based proxy re-encryption has following steps:

Fig. 1.   Overview of El Gamal-based Proxy Encryption

Schemes	Uni/Bi Directional	Security	Collusion Resistant
Blaze *et. al* [3]	$\leftrightarrow$	CPA	$\times$
Ivan and Dodis [15]	$\rightarrow$	CCA	$\times$
Ateniese *et al.* [2]	$\rightarrow$	CPA	$\sqrt{}$
Canetti and Hohenberger [6]	$\leftrightarrow$	CCA	$\times$
Deng *et al.* [9]	$\rightarrow$	CCA	$\times$
Shao *et al.* [18]	$\rightarrow$	CCA	$\times$
Sherman *et al.* [7]	$\rightarrow$	CCA	$\sqrt{}$

- Initialization: Similar to the initialization step in the El Gama encryption, we obtain a tuple (p, b, r) and a private key x.
- Key Generation: For each user i, the private key x is divided into 2 parts $x = x_{i1} + x_{i2}$. The user keeps x_{i1}, and the proxy keeps x_{i2}.
- Encryption: The user i encrypts his data with a random number k, $0 < k < p$, and send the ciphertext to the cloud, $\mathsf{UEnc}(m) = (r^k, r^{kx_{i1}}.m)$.
- Re-Encryption: The proxy gets the ciphertext from the cloud and re-encrypts and decrypts it with the corresponding proxy keys of user i and j.
    - Encryption: $\mathsf{PEnc} = (r^k, (r^k)^{x_{i2}}.r^{kx_{i1}}.m) = (r^k, r^{kx}.m)$
    - Decryption: $\mathsf{PDec} = (r^k, (r^k)^{-x_{j2}}.r^{kx}.m) = (r^k, r^k x - x_{j2}.m) = (r^k, r^k x_{j1}.m)$
- Decryption: The user j gets data from the proxy and decrypts it to get the original plain text, $\mathsf{UDec} = (r^k, (r^k)^{-x_{j1}}.r^{kx_{j1}}.m) = (r^k, m)$

C. Related work

In 1998, Blaze et. al [3] introduced the concept of proxy re-encryption (PRE) where a semi-trusted proxy can use a "re-encryption key" $RK_{A \rightarrow B}$ to transform a ciphertext corresponding to Alice's key into ciphertext for Bob's key, and the proxy cannot get the plain text. Although this encryption scheme is elegant and can against chosen plain text attack (CPA), it has a few limitations as follows:

- Bidirectional scheme: The proxy can compute $(RK_{A \rightarrow B})^{-1} = RK_{B \rightarrow A}$, which means that the proxy can re-encrypt Bob's messages under Alice's key, and Bob may not agree on this. An encryption scheme that allows the proxy can only convert in one direction is called unidirectional scheme.
- Collusion: If the proxy and Alice collude, they both can know Bob's secret key and vice versus.

Since the introduction of PRE by Blaze et. al., a number of different PRE schemes have been proposed with different security properties such as chosen ciphertext attack (CCA) security , collusion attack security.

In 2003, Ivan and Dodis [15] proposed a CCA security model for PRE. However, this scheme requires delegatees to pre-share a secret, which is undesirable in practice, for example Alice and Bob have no prior relationship and they do not want to communicate.

In 2005, Ateniese et al. [2] constructed a first unidirectional re-encryption without any required pre-sharing between parties, based on bilinear maps. Furthermore, this scheme is strong to collusion resistant in that the proxy is unable to collude with delegatees in order to expose the delegator's secret. The scheme's construction is efficient, semantically secure under the Bilinear DiffieHellman Inversion assumption [5].

In 2007, Canetti and Hohenberger [6] presented a CCA, bidirectional PRE scheme. This scheme is resistant to CCA attack that means ciphertexts remain indistinguishable even if the adversary can access to a re-encryption oracle and a decryption oracle. However, as their construction is bidirectional, there still a need to construct a CCA security PRE scheme in an unidirectional manner.

Recently, there are some publications on designing a CCA, unidirectional PRE scheme [9], [18], [7]. The properties of above schemes are summarized in Table I

III. THE FRAMEWORK

In this section, we introduce a secure framework in which every user has his secret key. However, any user is able to securely share his own data for all users in a same group without distributing his secret key. Moreover, if a user leaves the group, he no longer decrypt group's data even though he still can access the data. The framework is illustrated in Figure 2. There are four components in our proposed framework as we discuss following.

A. Framework Overview

1) Cloud Provider: Storing data on the cloud is provided as part of Infrastructure-as-a-Service (IaaS). It involves the storage of information on any number of virtual servers, by a third party (a cloud provider) rather than on servers owned and maintained by an enterprise itself. Storage of data in the cloud has several rewards including elasticity, back-up and universal access. With these advantages, more and more social network services such as Facebook, Twitter, etc. are moving to cloud to store their data that is generated by their users. However, due to its open access, the privacy and security of user's data becomes a major issue. Naturally, to enhance data privacy, users can encrypted their data before storing on cloud server. In our framework, the cloud provider's role is only supply the data storage for users and it is transparent to users and proxies.
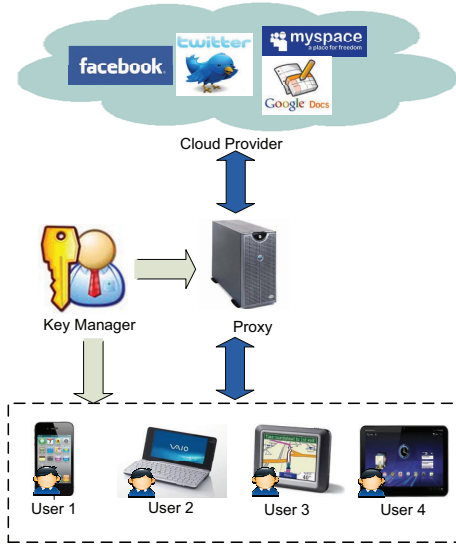
Fig. 2.   The Framework

2) Key Manager: Our framework uses a public-key proxy re-encryption scheme in which every user has his own key. To obtain this purpose, a key manager is needed to help distribute keys to all users as well as a proxy server. The Key Manager should be a third-trusted party. The tasks of the key manager are as follows.

- Generate a public-key encryption scheme with public key (p, r, h) and secret key x as in Section II-B2.
- Generate n pairs (x_{i1}, x_{i2}) so that $x_{i1} + x_{i2} = x$, and send x_{i1} to user i and (i, x_{i2}) to the proxy as shown in Table II, for all $i \in [1, n]$.

TABLE II
SECRET KEYS DISTRIBUTED TO USERS AND PROXY.

User key	x_{11}	x_{21}	...	x_{n1}
Proxy key	$(1, x_{12})$	$(2, x_{22})$	...	(n, x_{n2})

3) The Proxy: The proxy is a server standing between users and cloud provider. Its tasks include re-encryption and pre-decryption operations. An user i encrypts his message m with his key x_{i1} to send $c_i(m) = \mathsf{Enc}_{x_{i1}}(m)$ to the proxy. The proxy then uses the corresponding key x_{i2} to re-encrypt and sends $c(m) = \mathsf{Enc}_{x_{i2}}(c_i(m))$ to cloud storage.

On the other hand, if an user j wants to read the message m stored on the cloud storage. First, the proxy pre-decrypts the ciphertext $c(m)$ to get $c_j(m) = \mathsf{Dec}_{x_{j2}}(c(m))$ and sends it to user j. User j now uses his secret key to decrypt and gets the message: $m = \mathsf{Dec}_{x_{j1}}(c_j(m))$.

4) End-Users: Users in our framework use not only PCs but also mobile devices such as handphones, GPSes, etc., which allow them to upload and access social network data from anywhere. Users get their secret keys distributed by Key Manager. They then use the keys to encrypt their private data or decrypt data got from the proxy.

B. Keyword Search

Data stored on cloud storage is useless if users cannot search or query on it. Boneh et al. [4] proposed a public-key encryption scheme that helps users query on encrypted data. Nonetheless, in their method, all user use the same public key and only one user holds the secret key. In this section, we present a keyword encryption scheme to allow users be able to search on the encrypted data based on pre-defined keywords. The keyword encryption scheme is as follows.

- Initialization: Key Manager runs the Initialization step of ElGamal-based proxy pre-encryption in Section II-B2 to obtain (p, b, r, x). He chooses a hash function H, a random function f and a random key s for f. He then publicizes (p, b, r, H, f) and keeps x, s as the secret master key.
- Key Generation: Key Manager runs the Key Generation step of ElGamal-based proxy encryption in Section II-B2 to obtain (x_{i1}, x_{i2}). He distributes (x_{i1}, s) to user i and (i, x_{i2}) to the proxy.
- Keyword Encryption: User i chooses a random number k with $0 < k < p$. The keyword key is encrypted as: $\overline{c_1} = g^{k+\mu}$ where $\mu = f_s(key)$, $\overline{c_2} = \overline{c_1}^{x_{i1}}$, and $\overline{c_3} = H(b^k)$. User sends $\overline{c} \equiv (\overline{c_1}, \overline{c_2}, \overline{c_3})$ to the proxy
- Keyword Re-Encryption: The proxy calculates the cipertext of keyword key as: $c(key) = (c_1, c_2)$ where $c_1 = \overline{c_1}^{x_{i2}} \cdot \overline{c_2} = \overline{c_1}^{x_{i1}+xi2} = (g^{k+\mu})^x = b^{k+\mu}$ and $c_2 = \overline{c_3} = H(b^k)$.

The above keyword encryption scheme is used to check whether there is a match. Therefore, users can use it for pre-defined keywords search as well.

C. User Management

Users may register as a new member of a group or leaving the group. In this section, we propose several processes that help manage users in a social network group.

1) Registration: The main problem in a secure sharing data group is how to distribute users' secret keys. Our framework makes it simple by using a proxy re-encryption scheme. In a real group, the key manager may be the administrator of the group. When a user wants to join a group to share and access data, he sends a request to the key manager of the group. The key manager generates an user ID i and a pair (x_{i1}, x_{i2}). He distributes x_{i1} to user i and (i, x_{i2}) to the proxy. The user is now ready to share data or access data of other users in the group.

2) Revocation: Naturally, an user may want to leave the group or he may be removed from the group by the administrator. Thus, our framework supports user revocation in a very straightforward way. The key manager sends a request to the proxy to delete the proxy side key (i, x_{i2}). After the proxy side key is removed, the user cannot share or access data stored on the cloud storage anymore. It is clear that even the user still keeps his secret key and public key, he is unable to access the data since they proxy cannot pre-decryption encrypted data corresponding to user's key. If he would like to join the group

again, he needs to send a request of registration to the key manager.

3) Authentication: The framework allows the proxy and a user to establish a secure channel between them. Let see an authentication scheme as follows.

- User i sends a request of authentication to the proxy that contains his user ID i.
- The proxy retrieve user's proxy key x_{i2} based on user ID i and generate a random message M. The proxy encrypts M as $c(M) \equiv (g^r, h^r.g^{-r.x_{i2}}.M)$ and sends to the user.
- The user decrypts $c(M)$ as $g^{-r.x_{i1}}.h^r.g^{-r.x_{i2}}.M \equiv M$. He encrypts message $M-1$ as $c(M-1) \equiv (g^{r'}, h^{r'}.g^{-r'.x_{i1}}.(M-1))$ and sends to the proxy.
- The proxy decrypts $c(M-1)$ and checks if $M-1$ is correct.

Using above authentication scheme, the proxy can detect any unauthorized access to data. Only user with the key generated by the key manager can read the data store on cloud storage.

D. Security Analysis

From the description of the framework in the previous section, we address some potential threats as follows.

1) Does the proxy know user's data?: All users' data is encrypted with their own key before sending to the proxy. Therefore, if the proxy re-encryption scheme is semantically secure, the proxy cannot know any information from any user's data. The framework allows any users but the proxy can read the group data.

2) Collusion Attacks: The major threat in the framework is the collusion between the proxy and one of the users, even with a user who have left the group. If this collusion occurs, based on the keys distributed from the key manager, the proxy can disclose all other users' key. For instance, let say user j has left the group. He now colludes with the proxy to reveal others' key. He sends his key x_{j1} to the proxy. For any other user i, the proxy can easily reveal his key: $x_{i1} = x - x_{i2} = x_{j1} + x_{j2} - x_{i2}$ where x_{j2} and x_{i2} are distributed to the proxy in advance.

IV. CONCLUSION

In this paper, we have introduced a secure framework for data sharing among users of a group. Our framework provides the keyword search function that help users query on the group's data based on pre-defined keywords. The framework also supplies user management functions such as user registration or revocation for the administrator/key manager. In the framework, each user holds a secret key and thus, it allows the group administrator to easily revoke an user from the group. Since the keys are shared on both users and the proxy, once removed from the group, user is unable to access the group's data even though he still keeps his key. Moreover, the framework allows the proxy to efficiently authorize users via a simple scheme. In our future work, we will address the issue of collusion between the proxy and one of the users to reveal the other users's keys. Another important issue is the work load on the proxy which may suffer too many encryption or decryption operations.

REFERENCES

[1] Amazon. Building facebook applications on aws. http://aws.amazon.com/solutions/global-solution-providers/facebook.

[2] G. Ateniese, K. Fu, M. Green, and S. Hohenberger. Improved proxy re-encryption schemes with applications to secure distributed storage. *ACM Trans. Inf. Syst. Secur.*, 9(1):1–30, 2006.

[3] M. Blaze, G. Bleumer, and M. Strauss. Divertible protocols and atomic proxy cryptography. In *EUROCRYPT*. Springer-Verlag, 1998.

[4] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In *Proceedings of Eurocrypt 2004*, volume 3027, pages 506–522, 2004.

[5] D. Boneh and M. K. Franklin. Identity-based encryption from the weil pairing, 2001.

[6] R. Canetti and S. Hohenberger. Chosen-ciphertext secure proxy re-encryption, 2007.

[7] S. Chow, J. Weng, Y. Yang, and R. Deng. *Efficient Unidirectional Proxy Re-Encryption*, volume 6055 of *Lecture Notes in Computer Science*, pages 316–332. Springer Berlin / Heidelberg, 2010.

[8] D. Connor. Storage outsourcing on the rise. http://www.networkworld.com/news/2007/012207-storage-outsourcing-rises.html, 2007.

[9] R. H. Deng, J. Weng, S. Liu, and K. Chen. Chosen-ciphertext secure proxy re-encryption without pairings, 2008.

[10] Dropbox. Dropbox. http://www.dropbox.com.

[11] T. ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In *Proceedings of CRYPTO 84 on Advances in cryptology*, 1985.

[12] M. Feingold, M. Corzine, M. Wyden, and M. Nelson. Data mining moratorium act of 2003. U.S. Senate Bill (proposed), Jan 2003.

[13] S. L. Garfinkel. An evaluation of amazon's grid computing services: Ec2, s3 and sqs. Technical report.

[14] Google. What is google app engine? http://code.google.com/appengine/docs/whatisgoogleappengine.html.

[15] A. Ivan and Y. Dodis. Proxy cryptography revisited. In *Network and Distributed System Security Symposium*, 2003.

[16] Microsoft. Windows azure. http://www.microsoft.com/windowsazure.

[17] ScribeSoft. Scribe software expands cloud connectivity with release of web services adapter. http://www.scribesoft.com/PR022.asp.

[18] J. Shao and Z. Cao. Cca-secure proxy re-encryption without pairings, 2009.

[19] TheInfoPro. Theinfopro. http://www.theinfopro.com.

[20] Wikipedia. Facebook. http://en.wikipedia.org/wiki/Facebook.