

Selective Disclosure on Encrypted Documents

Hao Lei* and Dengguo Feng

State Key Laboratory of Information Security(SKLOIS), China
leiyokster@gmail.com

Abstract. With tackling the dilemma between the privacy concern and information utilization in mind, an efficient pairing-based instantiation of a new primitive, which we call Selective Disclosure scheme, is proposed in this paper. Selective Disclosure scheme allows the document issuer to distribute and publish the outsourced document in a secure way such that it achieves fine-grained authorized reading by selective parts in a document to different visitors and only one copy is needed. It is proved secure against fully adaptive adversaries in the random oracle model. The Selective Disclosure technique will be of use by embedding or integrating it into various word processors, e-mail, etc.

Keywords: Selective Disclosure, Privacy-Preserved Information Utilization, Plaintext Awareness Secure.

1 Introduction

With respect to all content in an outsourced document, it is reasonable that the content is selective disclosure to different visitors for the sake of document owner's privacy. That is, every visitor only has access to the information specified by the document owner, who totally controls and puts on different restrictions for different visitors. So, for the same document, the content read by different visitors is different. For example, in the case of a Blog document, a document visitor may be Blog owner's soul-mate, real friend, good friend, or just a simple friend. In this case, although they are all permissible visitors, it is reasonable that what a simple friend can read is different from what a soul-mate can.

In general, assume there are n kinds of permissible visitors related to a document, that is, n is the number of selective disclosure views stemming from this document. Once attempting to achieve above selective disclosure goals, a common approach appears below using known techniques (1) According to the dedicated content to be presented, the document owner creates n different copies from the original document. (2) For every copy, document owner selects n different keys, generates n different encrypted copies, and outsources them to the storage server. (3) The document owner distributes the corresponding decryption key to every matched permissible visitor. (4) Every permissible visitor can decrypt the encrypted dedicated copy and obtain the selective disclosure view.

* This is the extended abstract, June 3, 2011.

This way has two significant drawbacks in terms of efficiency, functionality and security analysis: (1) It is very troublesome for the document owner to generate n different dedicated copies and n different encrypted copies. (2) These n encrypted dedicated copies would require more storage space because of the redundant content among them.

Besides the above two applications, we find the selective disclosure also plays an important role in the scenario of outsourced storage [9], peer-to-peer storage systems [6,10], long-term archives[11], and web-service object stores [16], all of which share both information utilization and privacy concerns. In general, for the sake of data security and privacy, the outsourced document must be encrypted by the document owner before outsourcing to a third party storage provider.

1.1 Requirements, and Related Technologies

From above observations, the main security requirements of Selective Disclosure (SD for short) on outsourced documents can be phased as (1) It is the document owner that controls the content disclosed to different visitors, which means that the restrictions to every parts of the outsourced document are to be set by the document owner on his/her own, instead of a so-called trusted administrator, and SD guarantees the restrictions can be enforced correctly and strictly. This is termed as *document owner centric control property*. (2) Visitors with different permission can review different content, but they cannot read more content even if all of them collude against the document owner.

In the view of practical concerns such as functionality and performance, the following three requirements must also be taken into accounts (1) It admits any selective part (Paragraphs, Sections, Pages, etc.) contained in a document, and any subset of possible visitors, both of which are chosen ad hoc by the document owner. (2) It allows that every visitor can perform decryption independently without cooperation or any help from others. (3) Last but not the least, regardless of multiple views stemming from the original document, only one real copy of this document remains in storage and it incurs no additional storage costs.

Besides the above 5 requirements, a practical Selective Disclosure scheme must be provably secure without doubt.

The technologies that are closely related to Selective Disclosure are in three different areas, namely (1) *Traditional Access Control Approach*, (2) *Revocation and Broadcast Encryption*[3,15,14,12,5,4,7,13], and *Attribute-Based Encryption*[1,8]. The detailed introduction about related work as well as the difference between SD and them is listed in full version.

1.2 Key Idea, Challenges, and Our Contribution

The key idea to achieve the five requirements of SD at the same time is to aggregate all impermissible visitors for each part as a whole and use it to construct cipher text.

Our approach presents us with two challenges. First, we need to make sure that an impermissible visitor cannot do anything useful with his/her private key,

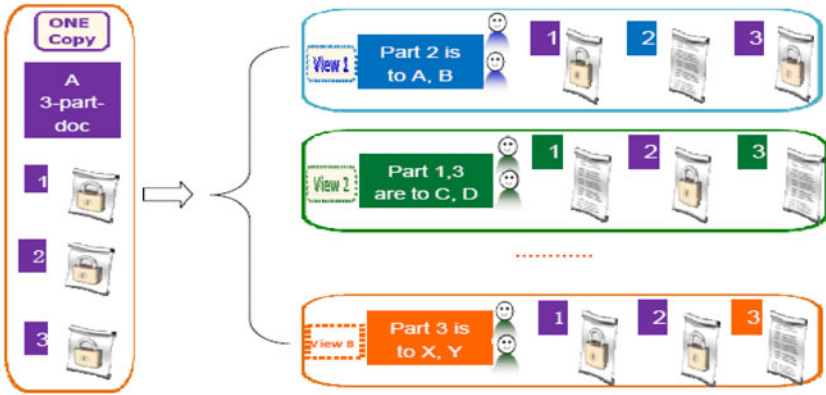


Fig. 1. SD enables eight views stemming from a three-part-document

even they collude against the document owner. Second, simply combining the ABE and revocation technologies is not sufficient, and we must ensure that not only the public and private keys, but also the ciphertext are of size independent of the number of permissible visitors.

With tackling the dilemma between the privacy concern and information utilization in mind, we propose an efficient pairing-based instantiation of a new primitive, which we call Selective Disclosure scheme. It allows the document issuer to distribute and publish the outsourced document in a secured way such that it achieves fine-grained authorized reading by selective parts in a document to different visitors and only one copy is needed. We then proved it is secure against fully adaptive adversaries in the random oracle model. To the best of our knowledge, there is no previous solution to enable selective disclosure on a document while incurs no additional storage consumption. The following Figure 1 presents a document with 3 parts which are selective disclosure to 8 different subsets of permissible visitors.

2 Syntax for SD Scheme

Before presenting our SD scheme we briefly review the definition as well as its security formulation for an SD scheme. If S is a set then $s \in_R S$ denotes the operation of picking an element s of S uniformly at random. We write $A(x, y, \dots)$ to indicate that A is an algorithm with inputs x, y, \dots and by $z \xleftarrow{R} A(x, y, \dots)$ we denote the operation of running probabilistic algorithm A with inputs x, y, \dots and letting z be the output. $z \leftarrow A(x, y, \dots)$ indicates that A is a deterministic algorithm.

2.1 Syntax for SD Scheme

Regarding to a part m contained in document, let Q denote all visitors who are represented by his/her public key pk_i , and $\overline{Q}_R \subset Q$ be a subset of impermissible visitors to m . A SD scheme consists of the following four algorithms:

$(Msk, Params) \stackrel{R}{\leftarrow} Setup(1^k)$. The setup algorithm, on input security parameter 1^k , outputs a master secret key Msk and public parameters $Params$.

$(PK_i, SK_i) \stackrel{R}{\leftarrow} KeyGen(Msk, Params, ID_i)$. The key generation algorithm takes as input master secret key Msk , public parameters $Params$ and user's identifier ID_i , outputs the corresponding public key PK_i and private key SK_i .

$C \stackrel{R}{\leftarrow} Enc(\overline{Q}_R, Params, m, SK_s)$. The encryption algorithm takes as input a document's portion $m \in \{0, 1\}^{k_0}$ together with public parameters $Params$, the owner's private key SK_s , and $\overline{Q}_R \subset Q$ which are impermissible visitors to m , outputs C which is the encryption of m for permissible visitors Q/\overline{Q}_R . The encryption algorithm is run by the document owner.

$m/\perp \leftarrow Dec(C, \overline{Q}_R, Params, SK_i, PK_i)$. The decryption algorithm takes as input the cipher text C together with public parameters $Params$, the permissible visitor private key SK_i and all the public keys of impermissible visitors \overline{Q}_R to m , outputs correct plaintext $m \in \{0, 1\}^{k_0}$. Otherwise, it returns \perp . This algorithm is run by anyone of the permissible visitors independently.

2.2 The IND-SD-CPA Security Game and Plaintext Awareness

The security for a SD scheme $\Pi = (Setup, KeyGen, Enc, Dec)$ is formulated by the following IND-SD-CPA experiment between an attacker A and a challenger B . Furthermore, we strengthen the standard definition of IND-CPA game by allowing the attacker A to issue chosen private key extraction queries.

Experiment $Exp_{A,\Pi}^{IND-SD-CPA-b}(k)$

$(Msk, Params) \stackrel{R}{\leftarrow} Setup(g_1, g_2, e(\cdot, \cdot))$.

$\forall ID_i \in \{0, 1\}^*, (pk_i, sk_i) \stackrel{R}{\leftarrow} KeyGen(Msk, Params, ID_i)$.

$Q = \{pk_1, \dots, pk_n\}$.

$(m_0, m_1, ID_c) \leftarrow A^{O_H, O_{sk}}(Params, Q)$.

$pk_c \leftarrow KeyGen(Msk, Params, ID_c)$.

$b \in_R \{0, 1\}; \overline{Q}_R = Q/\{(x_c, pk_c)\}; C^* \leftarrow Enc_{\overline{Q}_R}(m_b)$.

$b' \leftarrow A^{O_H, O_{sk}}(C^*, Params, Q)$.

return b'

Definition 1(IND-SD-CPA Secure)[2]. A SD scheme Π is secure against IND-CPA if for $k \in N$ and $b \in_R \{0, 1\}$, $Adv_{A,\Pi}^{IND-SD-CPA}(k) = |Pr[Exp_{A,\Pi}^{IND-SD-CPA-1}(k) = 1] - Pr[Exp_{A,\Pi}^{IND-SD-CPA-0}(k) = 1]|$ is negligible. The probability is over the random bits consumed by both the challenger B and adversary A .

Plaintext Awareness(PA) was defined in [2] and it formalizes an adversary's inability to create ciphertext without knowing its corresponding plaintext m . PA can be achieved through constructing a $\lambda(k)$ -Knowledge Extractor K on the basis of proved secure in the sense of IND-CPA, and PA implies SD is security against IND-CCA2 [2]. The following is a formal definition for PA and $\lambda(k)$ -Knowledge Extractor.

Let $\Pi = (Setup, KeyGen, Enc, Dec)$ be an encryption scheme, let B be an adversary, and let K be a knowledge extractor. For every $k \in N$ define:

$$Succ_{II,B,K}^{PA}(k) \stackrel{def}{=} \Pr[H \leftarrow Hash; (pk, sk) \leftarrow K(k); (hH, C, y) \leftarrow \text{runB}^{OH, Enc_{pk}^H}(pk) : K(hH, C, y, pk) = D_{sk}^H(y)] \geq \lambda(k)$$

Definition 2 ($\lambda(k)$ -Knowledge Extractor [2]). We say that K is a $\lambda(k)$ -extractor if K has running time polynomial in the length of its inputs and for every B , $Succ_{II,B,K}^{PA} \geq \lambda(k)$ where $1 - \lambda(k)$ is negligible and $y \notin C$, where C is the queried cipher text set.

Definition 3 (Plaintext Awareness Secure [2]). We say that II is secure in the sense of PA if II is secure in the sense of IND-CPA and there exists a $\lambda(k)$ -extractor K .

3 A CCA2 Secure SD Scheme and Its Application

SD scheme is a pairing-based cryptology methodology based on the following general decisional q-BDHI assumption. Let G_1, G_2 and G_T be cyclic groups with the same prime order p ($|p| = k$), where $k = k_0 + k_1$ is the security parameter. The parameter k_0 determines the size of plaintext to be encrypted, i.e., $m \in \{0, 1\}^{k_0}$. There exists an efficient computationally bilinear map $e : G_1 \times G_2 \rightarrow G_T$ with bilinearity and non-degeneracy properties, and a computable isomorphism $\psi : G_2 \rightarrow G_1$.

General Decisional q-BDHI Assumption is defined as follows.

Definition 4 (General Decisional q-BDHI Assumption). Taking a $(q + 4)$ -tuple $(x_c, g_1, g_2, g_2^\gamma, g_2^{\gamma^2}, \dots, g_2^{\gamma^q}, T) \in Z_p^* \times G_1 \times G_2^{q+1} \times G_T$ as input where $\gamma \in_R Z_p^*$, no P.P.T. adversary A has non-negligible advantage $\varepsilon(k)$ in distinguishing whether T is $e(g_1, g_2)^{1/(\gamma+x_c)}$ or a random in group G_T . That is, with respect to $|\Pr[A(x_c, g_1, g_2, g_2^\gamma, g_2^{\gamma^2}, \dots, g_2^{\gamma^q}, e(g_1, g_2)^{1/(\gamma+x_c)}) = 1] - \Pr[A(x_c, g_1, g_2, g_2^\gamma, g_2^{\gamma^2}, \dots, g_2^{\gamma^q}, T) = 1]|$, the advantage $\varepsilon(k)$ for any P.P.T. adversary A is negligible.

Let g_1 and g_2 be a generator of G_1 and G_2 respectively. Let $Q = \{pk_1, \dots, pk_{q-1}\}$ denotes all of the visitors (including document owner) in a SD scheme where $pk_i = (ID_i, x_i, B_i)$ and the ID_i is the identifier of visitor i . The IND-SD-CCA2 scheme is proposed in the following Table 1.

It is easy to see that the decryption algorithm is consistent. Indeed, if C is a valid cipher text, then we have

$$\begin{aligned} D_2 &= e(C_1, F^{\bar{Q}_R} \cup \{x_i\}) \cdot e(A_i, C_2) \cdot D_1 \\ &= e(u^t, F^{\bar{Q}_R} \cup \{x_i\}) \cdot e(A_i, w^t \cdot (F^{\bar{Q}_R})^t) \cdot e(u^t, B_i)^{-x_i} \\ &= e(h^{\gamma t}, (F^{\bar{Q}_R})^{\frac{1}{\gamma+x_i}}) \cdot e(h^{\frac{x_i}{\gamma+x_i}}, (F^{\bar{Q}_R})^t) \cdot e(h^{\frac{x_i}{\gamma+x_i}}, g_2^{\gamma t}) \cdot e(h^{\gamma t}, g_2^{\frac{1}{\gamma+x_i}})^{-x_i} \\ &= e(h, F^{\bar{Q}_R})^t \in G_T. \end{aligned}$$

Then $\frac{C_3}{D_2} = \frac{e(A_s, v)^t \cdot e(h, F^{\bar{Q}_R})^t}{e(h, F^{\bar{Q}_R})^t} = e(A_s, v)^t$, hence $\hat{M} = C_4 \oplus H_2(\frac{C_3}{D_2}) = C_4 \oplus H_2(V) = \hat{m} || \hat{s}$.

The algorithm for computing $F^{\bar{Q}_R} = g_2^{\frac{1}{(\gamma+x_1) \cdot (\gamma+x_2) \cdot \dots \cdot (\gamma+x_d)}}$ $\in G_2$ can be found in [14], i.e., to aggregate all impermissible visitors \bar{Q}_R on m . In addition, for efficiency consideration as well as avoiding the direct application of private key

Table 1. The IND-SD-CCA2 scheme

Setup (1^k)
$g_1, h \in_R G_1, g_2, v \in_R G_2, \gamma \in_R Z_p^*, u = h^\gamma \in G_1, w = g_2^\gamma \in G_2,$ $H: \{0, 1\}^* \rightarrow Z_p^*, H_1: \{0, 1\}^k \rightarrow Z_p^*, H_2: G_T \rightarrow \{0, 1\}^k, Msk \leftarrow \gamma,$ $Params \leftarrow (g_1, h, u = h^\gamma, g_2, w = g_2^\gamma, v, e(\cdot, \cdot), H, H_1, H_2).$ Output($Msk, Params$)
KeyGen ($Msk, Params, ID_i$)
$H(ID_i) = x_i, SK_i = A_i = h^{\frac{x_i}{\gamma+x_i}} \in G_1, B_i = g_2^{\frac{1}{\gamma+x_i}} \in G_2, PK_i = (ID_i, x_i, B_i).$ Output(SK_i, PK_i)
Enc ($m, Params, SK_s, \bar{Q}_R$)
$s \in_R \{0, 1\}^{k_1}, t = H_1(m s) \in Z_p^*, V = e(A_s, v)^t \in_R G_T,$ $F^{\bar{Q}_R} = g_2^{\frac{1}{(\gamma+x_1) \cdot (\gamma+x_2) \cdot \dots \cdot (\gamma+x_d)}} \in G_2,$ $C \leftarrow \{u^t, w^t \cdot (F^{\bar{Q}_R})^t, e(A_s, v)^t \cdot e(h, F^{\bar{Q}_R})^t, (m s) \oplus H_2(e(A_s, v)^t)\}.$ Output(C, \bar{Q}_R)
Dec ($C, \bar{Q}_R, Params, SK_i, PK_i$)
Parse C as $\{C_1, C_2, C_3, C_4\}$. If $PK_i \notin \bar{Q}_R: F^{\bar{Q}_R \cup \{x_i\}} = g_2^{\frac{1}{(\gamma+x_1) \cdot (\gamma+x_2) \cdot \dots \cdot (\gamma+x_v) \cdot (\gamma+x_i)}} = (F^{\bar{Q}_R})^{\frac{1}{\gamma+x_i}}.$ $D_1 = e(C_1, B_i)^{-x_i} = e(A_i, g_2^{\gamma t})^{-1} \in G_T,$ $D_2 = e(C_1, F^{\bar{Q}_R \cup \{x_i\}}) \cdot e(A_i, C_2) \cdot D_1 = e(h, F^{\bar{Q}_R})^t \in G_T,$ $\hat{M} = C_4 \oplus H_2(\frac{C_3}{D_2}) = C_4 \oplus H_2(e(A_s, v)^t), \hat{t} = H_1(\hat{M}) \in Z_p^*.$ If $C_1 \neq u^{\hat{t}}, C_2 \neq w^{\hat{t}} \cdot (F^{\bar{Q}_R})^{\hat{t}},$ output \perp . Output $m = [M]_{k_0}.$

A_s in practical use, the component $e(A_s, v)$ of C_3 can be pre-computed and one time pairing can be saved.

To testify the potential practical use of SD scheme, we show that SD scheme enables us to add selective disclosure property to the Microsoft Word in a secure manner. Firstly, the document owner highlights the dedicated part that desires to put restrictions on, then s/he selects the impermissible visitors from a list of all the potential visitors, which is analogous to the one of selecting recipients from the mail list. Secondly, the dedicated part will be encrypted according to SD scheme. Thus, every visitor only has access to the information content specified by the document owner, who totally restricts the information content and puts on different restriction for different visitors.

4 Security Proof for IND-SD-CCA2

Informally, the security of SD scheme is equivalent to the nonexistence of an adversary that is capable, within the confines of a certain game, of decrypting the cipher text on the condition that she/he is impermissible visitor.

By the definition 4.1 in [2], the security in the sense of IND-CPA and the existence of a knowledge extractor imply the security in the sense of Plaintext Awareness, which implies security against the adaptive chosen-ciphertext attack

(IND-CCA2) in virtue of the Theorem 4.2 in [2]. The detailed proof related to Theorem 1, Lemma 1 and Theorem 2 have been omitted because of space limit, and they are provided in the full version.

Firstly, the above SD scheme is IND-CPA secure according to Theorem 1.

Theorem 1. Let A be an adversary that has non-negligible advantage $\varepsilon(k)$ against the SD scheme in the sense of IND-SD-CPA. If the hash functions $H(\cdot)$, $H_1(\cdot)$ and $H_2(\cdot)$ are modeled as random oracles, and we let $q > 0$, $q_1 > 0$, $q_2 > 0$ and $q_{sk} > 0$ be the number of queries that A makes to $H(\cdot)$, $H_1(\cdot)$, $H_2(\cdot)$ and key generation oracle respectively. Then there is an algorithm B to solve the general decision q -BDHI problem in groups of order p with non-negligible advantage $\varepsilon(k)/2$.

Before conducting a knowledge extractor K , the following Lemma 1 elaborates $f_{u,v,w,h,\bar{Q}_R}(A_s, t)$ is injective. The partially trapdoor one-way function implied in the encryption function $\text{Encryption}(m, Params, SK_s, \bar{Q}_R)$ is defined as $f_{u,v,w,h,\bar{Q}_R}(A_s, t) \mapsto \{u^t, w^t \cdot (F^{\bar{Q}_R})^t, e(A_s, v)^t \cdot e(h, F^{\bar{Q}_R})^t\}$.

Lemma 1. The function $f_{u,v,w,h,\bar{Q}_R}(A_s, t)$ is injective.

Now we turn to construct a knowledge extractor K in Theorem 2.

Theorem 2. Let B be an adversary for PA . Then there is a knowledge $\lambda(k)$ -extractor K and hence the SD scheme is secure in the sense of PA , thus it is IND-CCA2.

5 Conclusion

In this paper, we proposed a secure Selective Disclosure Scheme that enables not only its content is selective disclosure to different document visitors, but also the dedicated information to every visitor is strict under the control of the document owner, while just requiring one real copy in storage and no additional storage consumption incurred.

Our work motivates two interesting open problems. The first is to find an efficient Selective Disclosure scheme in the case of a large number of impressive visitors. The second is to explore a Dynamic Selective Disclosure scheme, which comprises of the two requirements: (1) It allows new visitor added into existing permissible visitor set. Recall the proposed SD scheme only admits static visitor set, i.e., the impermissible visitor set must be determined prior to encryption. Otherwise, the new added member is able to read all the parts because he is not in any impressive visitor set. (2) It suits well to revoke permissible visitor, and/or make an impermissible visitor to be a permissible one, without constructing the cipher text from scratch.

Acknowledgments

Part of the work was done while the author was at High Privacy and Security group, NEC Labs., China. We are grateful to Ph.D Ke Zeng, and Ph.D Wenbin Chen for helpful discussions, and the anonymous reviewers for their comments.

References

1. Attrapadung, N., Imai, H.: Conjunctive Broadcast and Attribute-Based Encryption. In: Shacham, H., Waters, B. (eds.) Pairing 2009. LNCS, vol. 5671, pp. 248–265. Springer, Heidelberg (2009)
2. Bellare, M., Desai, A., Pointcheval, D., Rogaway, P.: Relations among Notions of Security for Public-key Encryption Schemes. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 26–45. Springer, Heidelberg (1998)
3. Fiat, A., Naor, M.: Broadcast Encryption. In: Stinson, D.R. (ed.) CRYPTO 1993. LNCS, vol. 773, pp. 480–491. Springer, Heidelberg (1994)
4. Goodrich, M.T., Sun, J.Z., Tamassia, R.: Efficient tree-based revocation in groups of low-state devices. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 511–527. Springer, Heidelberg (2004)
5. Halevy, D., Shamir, A.: The LSD Broadcast Encryption Scheme. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 47–62. Springer, Heidelberg (2002)
6. Kubiataowicz, J., Bindel, D., Chen, Y., et al.: Oceanstore: An architecture for global-scale persistent
7. Kurosawa, K., Desmedt, Y.: Optimum traitor tracing and asymmetric schemes. In: Nyberg, K. (ed.) EUROCRYPT 1998. LNCS, vol. 1403, pp. 145–157. Springer, Heidelberg (1998)
8. Li, J., Ren, K., Kim, K.: A²BE: Accountable Attribute-based Encryption for Abuse Free Access Control. Cryptology ePrint Archive, Report 2009/118 (2009)
9. Millstein, J.S., King, M., Morrison, Foerster, L.L.P.: Cloud Computing and Outsourcing: Is Data Lost in the Fog? (2009), <http://www.tawpi.org/uploadDocs/CloudComputingandOutsourcing.pdf>
10. Muthitacharoen, A.A., Morris, R., Gil, T.M., Chen, B.: Ivy: A read/write peer-to-peer file system. In: Proceedings of OSDI 2002, pp. 31–44 (2002)
11. Maniatis, P., Roussopoulos, M., Giuli, T., Rosenthal, D., Baker, M., Muliadi, Y.: The LOCKSS peer-to-peer digital preservation system. ACM Trans. on Computing Systems 23(1), 2–50 (2005)
12. Naor, D., Naor, M., Lotspiech, J.: Revocation and Tracing Schemes for Stateless Receivers. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 41–62. Springer, Heidelberg (2001)
13. Naor, M., Pinkas, B.: Efficient trace and revoke schemes. In: Frankel, Y. (ed.) FC 2000. LNCS, vol. 1962, pp. 1–20. Springer, Heidelberg (2001)
14. Delerablée, C., Paillier, P., Pointcheval, D.: Fully Collusion Secure Dynamic Broadcast Encryption with Constant-Size Ciphertexts or Decryption Keys. In: Takagi, T., Okamoto, T., Okamoto, E., Okamoto, T. (eds.) Pairing 2007. LNCS, vol. 4575, pp. 39–59. Springer, Heidelberg (2007)
15. Sahai, A., Waters, B.: Revocation Systems with Very Small Private Keys. Cryptology ePrint Archive, Report 2008/309 (2008)
16. Yumerefendi, A.Y., Chase, J.: Strong accountability for network storage. In: Proc. of FAST 2007, Trans. Storage, vol. 3(3) (2007)