Int. J. Advance. Soft Comput. Appl., Vol. 6, No. 1, March 2014
ISSN 2074-8523; Copyright © SCRG Publication, 2014

Secure Searchable Based Asymmetric Encryption in Cloud Computing

Majid Nateghizad, Majid Bakhtiari, Mohd Aizaini Maarof

Department of Computer Science, Faculty of Computing
Universiti Teknologi Malaysia
UTM Skudi, 81310 Johor, Malaysia
majid.nateghizad@gmail.com
bakhtiari@utm.my
aizaini@utm.my

Abstract

Cloud computing offers a wide range of services such as user demand network, operating systems, hardware, software and resources. These encouraging facilities and hassle free management of computing resources have attracted many users to outsource their data to untrusted servers. Cloud server may leak documents to unauthorized parties. Therefore, all documents need to be transmitted in ciphertext mode to keep user data confidential against an untrusted Cloud Service Provider (CSP). However, encrypting documents prevents user to search the outsourced documents directly. Regular encryption algorithms such as AES, RC4 and DES mechanisms have searching limitation; in which the whole ciphertext needs to be retrieved and then decrypt before search procedure can be performed. Recently, many advanced encryption techniques have been proposed to enable search capabilities for users. Generally, the keyword based search approach is used. This approach allows users to retrieve only those documents contain special keywords. However, searchable encryption algorithms suffer from security problems. This paper first explains Public key Encryption with Keyword Search (PEKS) algorithm and then proposes an improved secure searchable encryption algorithm based on Indistinguishability under Adaptive Chosen Ciphertext Attack (IND-CCA2). The proposed searchable encryption is mathematically proven secure and it has the ability to perform a search within encrypted data.

Keywords: *Cloud computing, PEKS, searchable encryption, keyword, security*

1 Introduction

Storage services over cloud are an essential part of cloud computing. It provides scalable data storage and computational services, and it minimizes the cost of resource management. In other words, customers can remotely outsource large amount of data and workloads to the cloud and benefit from unlimited computing resources and applications in the on-demand high-quality services. Relieving the burden for storage management, accessing to the data independent of locations, reducing in capital expenditure on hardware, software and staff are among the advantages of cloud computing brings to its users. Despite, privacy issue is critical factor that hinders the widespread adoption of cloud computing [1-4]. In other words, outsourcing data into clouds has many challenges such as security, privacy and control. For data outsourcing, the primary concern is data disclosure. Encrypting data before sending to cloud servers is a good approach to ensure confidentiality and prevent abuse of data by unauthorized user [5]. This approach can be found in the AES, RC4 and DES mechanisms .However, data encryption would limit people ability to handle their outsourced data; for example, user may want to retrieve only information contains set of specific keywords.

Searchable encryption is a cryptography primitive that enables users to search through outsourced encrypted data without exposing keywords to the untrusted server [6-8]. The area of searchable encryption has been known by The Defense Advanced Research Projects Agency (DARPA) as one of the technical advances for balancing between security and privacy.

Public key encryption is one of most common method in cryptography. In addition, this method is very useful to make a secure searchable encryption [9]. A user can encrypt any confidential data by a public key and then outsource the encrypted data to any cloud storage. Since, he is the only one holding the private key, no one but the authorized user has the ability to decrypt encrypted data. Having the ability of secure search through outsourced encrypted data should be the core feature of any searchable encryption and using public key encryption to make an algorithm has to fulfill that requirement. Secure search through outsourced encrypted data means authorized user is the only one can search for any keywords within outsourced data and unauthorized parties should not learn anything during search procedure.

Searching over outsourced encrypted data is suffering from three main problems [10, 11]: (1) Searching within outsourced encrypted data in cloud storage takes CPU and memory of both client side and server, which takes the availability cloud storage under risk. (2) Searching through outsourced encrypted data needs to be based specific procedures; it cannot be done simply by handing over the private key to the untrusted cloud storage, otherwise searching process threats confidentiality and integrity of outsourced data. (3) Encrypting, decrypting, and using hash function those are parts of any searchable encryption algorithm consume many CPU resources.

In this paper, we propose a new searchable encryption scheme and The contribution of our scheme can be summarized as follows: (1) The proposed searchable encryption supports keyword searching on encrypted data. It enables users to search desired word through ciphertext on CSP without revealing any information to the CSP. (2) The proposed encryption scheme is based on semantic security, which is resistant against adaptive chosen keyword attack (IND-CCA2). This paper is organized as follows: It begins by reviewing some related works and explains preliminaries. Later, it presents an outline of the proposed encryption scheme. Finally, this paper discusses the construction of the proposed encryption scheme and its security.

2 Related Work

Song *et al* [12] introduced the concept of searchable encryption. Searchable encryption can be categorized in two fields according to number of involved key: symmetric searchable encryption (SSE) and asymmetric searchable encryption (ASE).

2.1 Symmetric Searchable Encryption (SSE)

The first searchable symmetric encryption scheme proposed by Song [12]. In the Song's scheme, all the words within every document have to be encrypted in a double layer ciphertext form called inner layer and outer layer. Server strips the outer layer by using the trapdoor and checks the inner layer. In order to make the same verifiable inner layer structure trapdoor and ciphertext have to be generated by using the same keyword. The first index-based SSE proposed by Goh [13]. Goh's scheme is based on making a secure index of all the words in a document, which uses multiple different hash functions. The method of making that index and searching within that is called bloom filter, which is used in spread spectrum of applications in various areas. Then Curtmola [11] proposed another two inverted index-based SSEs, where its search time cost is O(1). Another SSE algorithm is proposed by Golle [4], which has the ability to search for conjunctive keywords. This algorithm is based on elliptic curve and suffered consuming high computational resources [14]. After that Tang [15] proposed a new SSE that supports phrase search.

2.2 Asymmetric Searchable Encryption (ASE)

The first asymmetric searchable encryption proposed by Boneh *et al* [9], public key encryption with keyword search. This algorithm is able to detect which encrypted outsourced document has a specific keyword without letting other parties such as CSP and unauthorized users to learn anything during search and retrieving process. This algorithm has the potential to be applied in real cloud computing with considering improvement in encryption and decryption algorithm,

which is not secure against adaptive chosen ciphertext attacks. Then Boneh *et al* [16] proposed an efficient and secure identity-based encryption algorithm. . The algorithm utilizes bilinear maps, which is secure for random oracle model. After that, many models constructed without random oracle but still satisfied the security. Recently, Boneh and Boyen [17] proposed a model in which random oracle is not used and model satisfied the security of data. However, the proposed algorithm is not practically possible and it is inefficient.

3 Preliminaries

In this section, this paper first introduces the basic definition of bilinear map and assumptions of Bilinear Diffie-Hellman (BDH), and then explains the PEKS scheme.

3.1 Assumptions

3.1.1 Bilinear Map

Assume G and G1 is a set of prime order, p in which there are computable bilinear map from G into G1.

i. Computable: e (g, h) ϵ G1 for given g, h ϵ G takes a polynomial time.

ii. Bilinear: any integers x , y ϵ [1,p] should satisfy $e(g^x, g^y)=e(g, g)^{xy}$

iii. Non-degenerate: e (g, g) ≠1 and also e(g, g) must be a generator of G1 if g be a random generator of G.

3.1.2 Decisional BDH assumption

Assume that challenger chooses four random values a, b, c and d ϵ Zp and other random binary value ß. Hence, there are two possibilities of ß, if ß=1 then it outputs are (g, A=g^a, B=g^b, C=g^c, z=$e(g,g)^{abc}$), otherwise the outputs are (g, A=g^a, B=g^b, C=g^c, z=$e(g,g)^z$. In this scenario an adversary has at least ε advantage to solve BDH problem where |PR[ß(g, g^a, g^b,g^c, $e(g,g)^{abc}$=1] - PR[ß(g, g^a,g^b, g^c, $e(g,g)^z$)=1]| ≥ 2ε

3.1.3 Computational BDH assumption

Randomly chosen a, b ϵ Z_p outputs (g, A =g^a, B=g^b).It is about advantages of adversary A in attempting to output [g^{ab}ϵ G is ε if PR [ß(g, g^a, g^b)=g^{ab}] ≥ ε.

3.2 PEKS algorithm

In PEKS algorithm, there are three parties called sender, receiver, and server. The "sender" is the one who creates and outsource ciphertext, which is called PEKS Ciphertext. The process of searching on outsourced PEKS ciphertext upon given trapdoor takes place in server. Finally, the receiver is the party that generates a trapdoor for any desire keyword to be searched within outsourced encrypted data in server. Fig.1 shows the process of PEKS, which is explained in more details as follow:

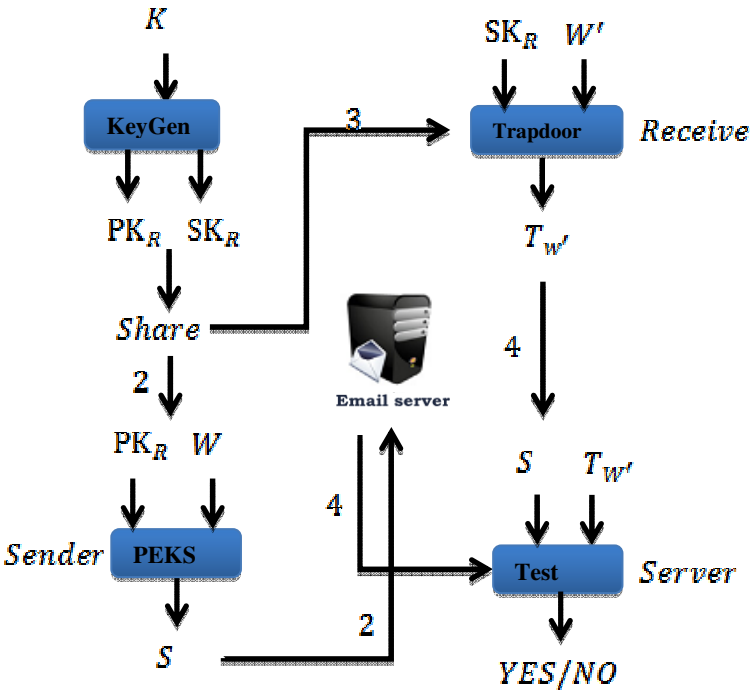

Figure 1: PEKS algorithm

A Receiver Key Generation Algorithm $\mathbf{KeyGen_{Receiver}}(k)$: It takes a security parameter K ϵ N as input and generates a public and private key $(\mathbf{SK}_R, \mathbf{PK}_R)$ of the receiver.

PEKS Algorithm $(\mathbf{PK}_R, W)$: This algorithm generates a PEKS ciphertext S, which is a searchable encryption of given word W and it uses receiver's public key PK_R as inputs for encryption function. According to inputs and output if this function the equation is $\mathbf{S = PEKS}\,(\mathbf{PK}_R, W)$.

A Trapdoor Generation Algorithm Trapdoor $(\mathbf{SK}_R, W')$: This function generates a trapdoor $T_{W'}$ of given word W', which is the word that receiver wants to search within outsourced encrypted data in cloud storage, and receiver's private key $\mathbf{SK}_R$. The server uses trapdoor $T_{W'}$ to perform search function.

A Test Function Test $(T_{W'}, S)$: This function takes receiver's trapdoor $T_{W'}$ and a PEKS ciphertext S from server, which is equal to $\mathbf{PEKS}\,(\mathbf{PK}_R, W)$ and then returns "YES" if $W = W'$ otherwise it returns "NO".

4 Ciphertext Indistinguishability

There are four categories of ciphertext indistinguishability and they are indistinguishability against Known Plaintext Attack (KPA), Chosen Plaintext Attack (CPA), Non-Adaptive Chosen Ciphertext Attack (CCA), and Adaptive Chosen Ciphertext Attack (CCA2).

4.1 Known plaintext attack

Known plaintext attack is a scenario in which the attacker has access to pairs of plaintexts and their corresponding ciphertexts. This attack considered as a high practical attack, especially if the number of pairs are not too many. Also, known plaintext attack is more practical than chosen plaintext attack and chosen ciphertext attack. Probable word method, which is a popular technique for classical simple substitution or transposition ciphers is an example of known plaintext attack.

4.2 Chosen plaintext attack

In this type of attack, it assumes that the attacker has the capability to choose any plaintext to be encrypted and then get the related ciphertext from the server. The goal of this attack is to reduce the security of encryption scheme by trying to retrieve the secret key from multiple pair of plaintext and ciphertext. At the first, this method of attack seems to be unrealistic because it is difficult to have human cryptographer to make decrypt that amount of plaintext are sent from the attacker. However, chosen plaintext attack is very feasible to happen because of the fact that modern cryptography methods are implemented in a hardware or software for spread spectrum of applications, which makes this type of attack easier. Although, this type of attack is considered less practical than known plaintext attack, still it is known as a dangerous attack.

4.3 Chosen ciphertext attack

Chosen ciphertext attack is an attack type in which the attacker has the ability to obtain corresponding plaintext from server of any chosen ciphertext by him. This scenario is almost the same as chosen plaintext attack but the attacker has access to decryption function, instead of encryption function. In public key cryptography, chosen ciphertext attack is potentially dangerous because chosen ciphertext is always available due to publicly known encryption key. According to this issue, RSA public-key encryption is not secure against adaptive chosen ciphertext attack. In addition, El-Gamal cryptosystem is vulnerable and insecure against chosen ciphertext attack, even though it is secure against chosen plaintext attack. Although the possibility of chosen ciphertext attack is less than chosen plaintext

attack and it is less practical, but there is no direct correspondence between chosen ciphertext and chosen plaintext attack complexities.

4.4 Adaptive chosen ciphertext attack

In adaptive chosen ciphertext attack, attacker uses decryption function of sever to decrypted prepared ciphertexts, which are generated via publicly encryption key. This attack type is based previous outcome of decryption function. In the non-adaptive definition, the adversary is allowed to query this oracle only until it receives the challenge ciphertext. In the adaptive definition, the adversary may continue to query the decryption oracle even after it has received a challenge ciphertext, with the caveat that it may not pass the challenge ciphertext for decryption. This scenario is more powerful and less realistic rather than none-adaptive chosen ciphertext attack. However, this attack can be quite practical in public encryption method. Bleichenbacher [18] proved that an adoptive chosen ciphertext attack can be carried out when only partial information about corresponding message is leaked.

5 Outline of the Proposed Searchable Encryption Algorithm

Assume that a user needs to outsource critical data in CSP. Those data contains a

series of messages m ϵM. In order to make this encryption searchable, we need to send encrypted keywords to the CSP along with encrypted data. For encrypting data and keywords, our proposed algorithm uses public key A_{pub} and for decrypting information or generating trapdoor, it uses two different private keys.

KeyGen(s): Given security parameter s and computes A_{pub} and A_{priv} as public/private keys.

Encryption (A_{pub}, m): Given public key A_{pub} and message M, encrypt the message M using A_{pub}

SEncryption (A_{pub}, W): Encrypt given word W in searchable form by using public key A_{pub}.

Trapdoor (A_{priv},W): Takes private key A_{priv} and word W to generate trapdoor T_W.

Test (A_{pub}, S, T_W): Search within searchable encrypted data S=PEKS (A_{pub}, W') in

CSP for a given trapdoor T_W and word W. It outputs yes if W= W' meaning this message containing word W.

Decryption (A_{priv}, E): it takes private key A_{priv} and encrypted message E to recover plain text m $\boldsymbol{\epsilon M}$.

6 Construction of the Proposed Searchable Encryption Algorithm

The proposed schema is based on bilinear map. Assume that we have a function that generates the required parameters G, G1, α, g, g_2 and $\boldsymbol{u'}$ where G and G1 are two groups of prime order *p* that a bilinear map e: G × G→G1 must satisfied. Meanwhile *h* is another computed variable based on *g* in which it takes a polynomial time to compute e(*g*,*h*) ϵ G1, e($\boldsymbol{g^x}$,$\boldsymbol{g^y}$) = $\boldsymbol{e(g,g)^{xy}}$ and *g* is a random generator of G which e(g,g) is a generator of G1. In the proposed algorithm two hash functions are used $\boldsymbol{H_1}$: $\boldsymbol{\{0,1\}^*}$ → G and $\boldsymbol{H_2}$: G1→ $\boldsymbol{\{0,1\}^{(log\ p)}}$.

KeyGen: Given that security parameter settles the size *p* of the groups G and G1. The proposed algorithm takes random values α,r ϵ Z_p*, picks a random generator g of G, chooses g_2, randomly pick u' in G. This function also generates a random *n*-length vector of U = (u_i) in which the elements of U are randomly picked from G. KeyGen function outputs A_{pub} = [g, h=g^{α}] and A_{priv}= α. Since the proposed algorithm is identity based, *v* is an *n* bits that stands for an identity and V_is the i^{th} bit of *v* where $i \subseteq \{1,\ldots, n\}$ in which $v_i = 1$. Finally, KeyGen outputs a private key to identify *v* as follows: $d_1 = g_2^{\alpha}(u' \prod_{i\in v} u_i)^r$ and $d_2 = g^r$, $d_v = (d_1, d_2)$ or $d_v = (g_2^{\alpha}(u' \prod_{i\in v} u_i)^r, g^r)$.

Encryption: In this step the message M ϵ G1 is encrypted for an identity by assuming that t ϵ Z_p is a random number. The generated encrypted message is as follows:
There are three parameters C_1, C_2, C_3 that will be sent to CSP. $C_1 = e(g_1, g_2)^t M$, $C_2 = g^t$, $C_3 = (u' \prod_{i\in v} u_i)^t$ and finally $C = (C_1, C_2, C_3)$ or $C = e(g_1, g_2)^t M, g^t, (u' \prod_{i\in v} u_i)^t)$.

SEncryption (A_{pub}, W): This function outputs encrypted W in such a way that is searchable in which public key is used for encryption and private key is used to perform a search within encrypted data. This process has two steps, first computes t = e (H_1 (W), h^r) ϵ G2 and then outputs [g^r, H_2 (t)].

Trapdoor (A_{priv}, W): This function encrypts a keyword with a different key in order to perform a search within encrypted data to find out which outsourced encrypted document contains the desired word W. Since the proposed algorithm is based on public key searchable encryption, user A encrypts tokenized keywords with a public key and uses private key to perform search.
Test (A_{pub}, S, $T_{W'}$): Gives A_{pub}, S and $T_{W'}$ and outputs 'yes' or 'no' where S = SEncryption(A_{pub}, W). In this step CSP searches $T_{W'}$ within outsourced encrypted data made by SEncryption() to find a match. The searching process is as follows: Let S= [A|B] where A=$g^r$and B= H_2 (t). Test if H_2 (e ($T_{W'}$, A)) =B, if so function outputs 'YES', otherwise outputs 'NO'.

Decryption: In the encryption process, three parameters are sent to CSP. They are $e(g_1, g_2)^t M, g^t$ and$(u' \prod_{i \in v} u_i)^t$, assume C_1=$e(g_1, g_2)^t M$, C_2=g^t and C_3=$(u' \prod_{i \in v} u_i)^t$

M is equal to

$$C_1 \frac{e(d_2, C_3)}{e(d_1, C_2)} \quad (1)$$

where it can be computed as:

$$e(g_1, g_2)^t M)(\frac{e(g^r, (u' \prod_{i \in v} u_i)^t)}{e(g_2^\alpha, (u' \prod_{i \in v} u_i)^r, g^t)} \quad (2)$$

Where

$$\frac{e(g^r, (u' \prod_{i \in v} u_i)^t)}{e(g_2^\alpha, (u' \prod_{i \in v} u_i)^r, g^t)} \quad (3)$$

is equal to

$$\frac{e(g, (u' \prod_{i \in v} u_i)^{rt})}{e(g_1, g_2)^t e((u' \prod_{i \in v} u_i)^{rt}, g)} \quad (4)$$

7 Security of the Proposed Algorithm

The encryption and decryption method that is used in this paper is based on the Waters IBE scheme [19], which is the most efficient IND-CCA2. In addition, the security of this algorithm is proved by reducing it to the decisional bilinear Diffie-Hellman problem. This encryption scheme is (t, q, ε) semantically secure

assuming (t + o(ε^{-2} ln(ε^{-1})λ^{-1} ln(λ^{-1})), ε/32(n+1)q) where λ=1/8(n+1)q if all t times adversaries making at most q private key queries have at most an ε in breaking IBE scheme [19].

8 Future Work

The proposed algorithm on this paper is based on practical IND-CCA2 security; however, this algorithm suffers from high search time cost, which is linear. It is possible to merge the Waters IBE [19] algorithm with the index based searchable encryption proposed by Curtmula [11]. Our future work is to investigate other search methods and propose a practical index-based secure searchable encryption algorithm that satisfies a reasonable performance concerning search time cost and security.

9 Conclusion

Nowadays CSPs offer many services in cloud that benefit people and organizations to reduce their Business management cost. However, using cloud storage brings some privacy issues. In order to provide confidentiality to the information in cloud, users need to encrypt their documents. This solution satisfies privacy of users' outsourced data in cloud but it reduces their control on their outsourced data. This drawback has motivated researchers to propose searchable encryption methods, which enables users to perform encrypted search queries on the server side. This paper proposed a new searchable encryption algorithm, which has proven to improve the security of searchable encryption to IND-CCA2.

ACKNOWLEDGEMENTS.
We would like to thank the anonymous reviewers for their useful comments and Dr. Imran Ghani for his help to improve this research paper.

References

[1] T. Mather, S. Kumaraswamy, and S. Latif. 2009. Cloud security and privacy: an enterprise perspective on risks and compliance: O'Reilly.

[2] S. Pearson and A. Benameur. 2010. "Privacy, security and trust issues arising from cloud computing," in *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pp. 693-702.

[3] M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou. "Security and privacy in cloud computing: A survey," in *Semantics Knowledge and Grid (SKG), 2010 Sixth International Conference on*, pp. 105-112.

[4] S. Subashini and V. Kavitha. "A survey on security issues in service delivery models of cloud computing," *Journal of Network and Computer Applications,* vol. 34, pp. 1-11.

[5] S. Zhang, S. Zhang, X. Chen, and X. Huo. "Cloud computing research and development trend," in *Future Networks, 2010. ICFN'10. Second International Conference on*, pp. 93-97.

[6] C. Wang, K. Ren, S. Yu, and K. M. R. Urs. "Achieving usable and privacy-assured similarity search over outsourced cloud data," in *INFOCOM, 2012 Proceedings IEEE*, pp. 451-459.

[7] M. Li, S. Yu, N. Cao, and W. Lou. "Authorized private keyword search over encrypted data in cloud computing," in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, pp. 383-392.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou. "Secure ranked keyword search over encrypted cloud data," in *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*, pp. 253-262.

[9] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano."Public key encryption with keyword search," in *Advances in Cryptology-Eurocrypt 2004*, pp. 506-522.

[10] M. Bellare, A. Boldyreva, and A. O'Neill. "Deterministic and efficiently searchable encryption," in *Advances in Cryptology-CRYPTO 2007*, ed: Springer, pp. 535-552.

[11] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security*, pp. 79-88.

[12] D. X. Song, D. Wagner, and A. Perrig. "Practical techniques for searches on encrypted data," in *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*, pp. 44-55.

[13] E.-J. Goh. 2003. "Secure Indexes," *IACR Cryptology ePrint Archive,* vol. 2003, p. 216.

[14] H. N. Lu, D. W. Gu, C. Y. Jin, and Y. Q. Tang. "Reducing Extra Storage in Searchable Symmetric Encryption Scheme," *2012 Ieee 4th International Conference on Cloud Computing Technology and Science (Cloudcom).*

[15] Y. Tang, D. Gu, N. Ding, and H. Lu. "Phrase Search over Encrypted Data with Symmetric Encryption Scheme," in *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*, pp. 471-480.

[16] D. Boneh and M. Franklin. "Identity-based encryption from the Weil pairing," in *Advances in Cryptology—CRYPTO 2001*, pp. 213-229.

[17] D. Boneh and X. Boyen."Efficient Selective Identity-Based Encryption Without Random Oracles," *Journal of Cryptology,* vol. 24, pp. 659-693, Oct.

[18] D. Bleichenbacher. "Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS# 1," in *Advances in Cryptology—CRYPTO'98*, pp. 1-12.

[19] B. Waters. 2005. "Efficient identity-based encryption without random oracles," in *Advances in Cryptology–EUROCRYPT 2005*, ed: Springer, pp. 114-127.