On the Security of the Identity-based Encryption based on DHIES from ASIACCS 2010

Willy Susilo*
Center for Computer and Information Security Research (CCISR)
School of Computer Science and Software Engineering
University of Wollongong, Australia
wsusilo@uow.edu.au

Joonsang Baek
Cryptography and Security Department
Institute for Infocomm Research (I2R) Singapore
jsbaek@i2r.a-star.edu.sg

ABSTRACT

In ASIACCS 2010, Chen, Charlemagne, Guan, Hu and Chen proposed an interesting construction of identity-based encryption based on DHIES, whose key extraction algorithm makes use of the multivariate quadratic equation. They proved that their scheme is selective-ID secure against chosen ciphertext attack, i.e. secure in the sense of `IND-sID-CCA`. Unfortunately, in this paper, we demonstrate that Chen et al.'s scheme is insecure in the sense of `IND-sID-CCA` by showing that the private key extraction algorithm of their scheme can be exploited to apply XL algorithm, which is to solve the multivariate quadratic (MQ) problem (under certain conditions).

Keywords

Identity-based Encryption, multivariate quadratic, XL

1. INTRODUCTION

In 1984, Shamir [17] introduced the notion of identity-based cryptography to simplify key management by avoiding the use of digital certificates by letting a public key be publicly derivable from identities, which can be any arbitrary string, such as an email address or a telephone number. More precisely, in order to generate a corresponding private key associated with a given identity, a trusted Private Key Generator (PKG) must compute the private key by using the knowledge of a master secret key. This paradigm avoids the complicated key management problems arising in traditional public key infrastructures as it eliminates the need for certificates and some of the problems associated with them. Hence, identity-based cryptography supports lightweight environment, in comparison to the traditional public key infrastructure. However, the dependence on a PKG who uses a system-wide master key to generate private keys inevitably introduces the inherent problem in identity-based cryptosystems, namely the key escrow problem.

In an identity-based encryption, an encryptor is only required to know who is allowed to decrypt a ciphertext. It is up to the decryptor to obtain the required key from the trusted PKG. In other words, an identity-based cryptosystem allows an encryptor to encrypt a message directed to a decryptor even *prior to* the decryptor retrieving the associated private key from the PKG. This "feature" enables interesting applications, such as access control that is defined by a logical formulae of conjunctions and disjunctions [19].

In [17], Shamir proposed a concrete construction of identity-based signature, but the construction of identity-based encryption remained as an open problem. Subsequently, two similar identity-based encryption based on pairings were independently proposed by Sakai, Ohgishi and Kasahara [16] and by Boneh and Franklin [7], followed afterwards by a completely different scheme by Cocks [10]. Since 2000, many improvements have been attained to the original scheme proposed by Boneh and Franklin, and new techniques have also been proposed (e.g. [14, 6, 5, 20]). For a complete overview of the improvements on identity-based encryption, we refer the readers to [8]. Interestingly, most of the schemes incorporate in one way or another upon the notion of bilinear pairings.

A more recent development of identity-based encryption schemes has been made using lattice-based cryptography [12, 15]. Lattice-based cryptography has attracted many attentions in the cryptographic community since it has promised a futuristic solution towards combating quantum computers ability [18], and hence the notion of post-quantum cryptography [4]. Additionally, code-based cryptography, hash-based cryptography and multi-variate cryptography have also been identified as the potential primitives in the post-quantum cryptography era [4].

Very recently, in ASIACCS 2010, Chen, Charlemagne, Guan, Hu and Chen [9] proposed an interesting construction of identity-based encryption based on the DHIES (Diffie-Hellman Integrated Encryption) [3, 1] with multivariate quadratic key combination structure and bilinear maps. The scheme is very efficient and it requires only one pairing computation during public key generation, and there is no special hash function required. Interestingly, they also provided a brief of security analysis based on the multivariate quadratic prob-

*This work is supported by ARC Future Fellowship FT0991397.

ASIACCS'11 March, 22–24, 2011, Hong Kong, China.

lem. Furthermore, the scheme is proven selective ID-secure [5] against chosen ciphertext attack, i.e. IND-sID-CCA secure, in the random oracle model assuming DHIES is chosen ciphertext secure. Moreover, they also claimed that their extract algorithm constitutes a short signature scheme, which is of an independent interest.

OUR CONTRIBUTION. In this paper, we provide a cryptanalysis of Chen et al.'s scheme [9]. In contrast to the claimed result in [9], we show that the proposed identity-based encryption scheme is not IND-sID-CCA secure. The insecurity arises from the KeyGen algorithm, which uses the multivariate quadratic equation to derive private keys. We show that due to the flexibility in making a number of private key extraction queries, the KeyGen algorithm can successfully exploited by the method proposed in [11]. We envision that the only fix for the scheme to be secure is to have it an unreasonably large public parameter.

ROAD-MAP. The rest of this paper is organized as follows. In Sec. 2, we recall some background that will be required throughout this paper. In Sec. 3, we provide a review on Chen et al.'s identity-based encryption scheme [9] from ASIACCS 2010. In Sec. 4, we demonstrate that the scheme is *not* IND-sID-CCA secure. Sec. 5 concludes the paper.

2. PRELIMINARIES

2.1 Admissible Bilinear Maps

Let k be a security parameter and q be a k-bit prime number. Let us consider groups $\mathbb{G}_1$ and $\mathbb{G}_2$ of the same prime order q. A bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ satisfies the following properties:

- Bilinearity: $\forall P, Q \in \mathbb{G}_1, \forall a, b \in \mathbb{Z}_q^*$, we have $e(aP, bQ) = e(P, Q)^{ab}$.
- Non-degeneracy: $\forall P \in \mathbb{G}_1$, $e(P, P) \neq 1$.
- Computability: $\forall P, Q \in \mathbb{G}_1$, $e(P, Q)$ can be efficiently computed.

As demonstrated in [7], such non-degenerate admissible maps over cyclic groups can be obtained from the Weil or the Tate pairing over algebraic curves.

2.2 Identity-based Encryption

We review the formal definition of identity-based encryption introduced in [7].

SETUP: is a probabilistic algorithm run by a private key generator (PKG) that takes as input a security parameter to output a public/private key pair for the PKG, denoted by (P_{pub}, mk).
KEYGEN: is a key generation algorithm run by the PKG on input a master key mk and a user's identity ID to output the user's private key d_{ID}.
ENCRYPT: is a probabilistic algorithm that takes as input a plaintext M, a recipient's identity ID and the PKG's public key P_{pub} to output a ciphertext C.
DECRYPT: is a deterministic decryption algorithm that takes as input a ciphertext C and the private decryption key d_{ID} to return a plaintext M or a symbol REJECT if C is an invalid ciphertext.

2.3 Security Notions

In the following, we review the definition of selective identity, adaptively chosen ciphertext security of an identity-based encryption scheme defined in [9].

DEFINITION 1. *An identity-based encryption scheme is said to be selective identity, adaptively chosen ciphertext secure (IND-sID-CCA) if no PPT adversary $\mathcal{A}$ has a non-negligible advantage in the following game.*

Init. *A outputs an identity ID_{ch} which it wishes to be challenged on.*

Setup. *The challenger runs the* SETUP *algorithm. It provides the adversary the resulting system parameters* PARAMS. *It keeps the* MASTER-KEY *to itself.*

Phase 1. *A issues queries $q_1, \cdots, q_m$ where query q_i is one of the following*

- *Extraction query (ID_i) where $\mathsf{ID}_i \neq \mathsf{ID}_{ch}$. The challenger responds by running the algorithm* KEYGEN *to generate the private key d_i corresponding to the identity ID_i. It sends d_i to the adversary.*
- *Decryption query (ID_i, C_i). The challenger responds by running the algorithm* KEYGEN *to generate the private key d_i corresponding to ID_i. Then, it runs the algorithm* DECRYPT *to decrypt the ciphertext C_i using the private key d_i. It sends the resulting plaintext to the adversary.*

Challenge. *Once the adversary decides that Phase 1 is over, it outputs two equal length plaintext M_0, M_1 on which it wishes to be challenged. The challenger picks a bit $c \in \{0, 1\}$ and sets the challenge ciphertext C =*ENCRYPT*(*PARAMS*, ID_c, M_c). It sends C as the challenge to the adversary.*

Phase 2. *The adversary issues additional queries $q_{m+1}, \cdots, q_n$ where q_i is one of the following*

- *Private key query $\mathsf{ID}_i \neq \mathsf{ID}$. The challenger responds the same as in Phase 1.*
- *Decryption query $(\mathsf{ID}_i, C_i) \neq (\mathsf{ID}_{ch}, C)$. The challenger responds the same as in Phase 1.*

These queries may be asked adaptively as in Phase 1.

Guess. *Finally, the adversary outputs a guess $c' \in \{0, 1\}$ and wins if $c = c'$.*

2.4 Multivariate Quadratic Problem

The multivariate quadratic (MQ) problem can be defined as follows [21]. Let $P_1, \cdots, P_m$ be m polynomials of n variables over $\mathbb{F}_q$, each of which has the form

$$P_t(x_1, \cdots, x_n) = \sum_{1 \leq i \leq j \leq n} \beta_{ij}^{(t)} x_i x_j + \sum_{i=1}^{n} \alpha_i^{(t)} x_i + \gamma^{(t)}$$

where $\beta_{ij}^{(t)}, \alpha_i^{(t)}, \gamma^{(t)} \in \mathbb{F}_q$ for all $1 \leq t \leq m$. The MQ problem, denoted as $\mathcal{MQ}(q, n, m)$, is the problem of solving for indeterminate $x_i \in \mathbb{F}_q$ of the random system of m Multivariate Quadratic equations $y_t = P_t(x_1, \cdots, x_n)$ for $1 \leq t \leq m$.

3. REVIEW ON CHEN ET AL.'S IDENTITY-BASED ENCRYPTION

The essence of Chen et al.'s scheme [9] is in the KeyGen algorithm. In order to generate a key pair for a given ID, the KeyGen algorithm first maps ID to a set S of indices, then the PKG computes the linear combination of some secret elements related to S as the private key, while everyone can compute the corresponding combination of the public elements related to S as the public key. The idea seems straightforward, but Chen et al. observed that the linear key combination structure will make the construction vulnerable to collusion attack. Therefore, they suggested to use a non-linear key combination structure [9]. The resulting scheme is based on DHIES with multivariate quadratic combination structure. The complete description of the scheme is as follows. In the following description, a symmetric key scheme $\mathsf{SE} = (\mathcal{K}_s, \mathcal{E}_s, \mathcal{D}_s)$ is used, where $\mathcal{E}_s(\mathtt{k}, \cdot)$ and $\mathcal{D}_s(\mathtt{k}, \cdot)$ denote symmetric encryption and decryption under the key $\mathtt{k}$, respectively.

Setup. Given a security parameter $k \in \mathbb{Z}^+$, and the PKG scale ℓ, the algorithm works as follows.

- Run $\mathcal{G}$ on input k to generate a prime q, two groups $\mathbb{G}_1 =< P >, \mathbb{G}_2 =< g >$ of order q, and an admissible bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$. Let $g \in \mathbb{G}_2$ be $e(P, P)$, which is the generator of $\mathbb{G}_2$.
- Generate an ℓ-dimension secret vector $\mathbb{SV} = (d_1, \cdots, d_\ell)$, where d_i is randomly selected from $\mathbb{Z}_q^*$.
- Generate the corresponding ℓ-dimension public vector $\mathbb{PV} = \mathbb{SV} \cdot P = (U_1, \cdots, U_\ell) = (d_1 P, \cdots, d_\ell P)$.
- Let $H_0 : \{0,1\}^* \to \{s_1, \cdots, s_t\} \subseteq \{1, \cdots, \ell\}$ be an identity mapping function, which maps an arbitrary identity string to a t-size subset of $\{1, \cdots, \ell\}$.
- Construct an identity map function $H_1 : \{0,1\}^* \to \mathbb{G}_2$ based on H_0.
- Choose a pseudorandom number generator RNG over $\mathbb{F}_q$, which takes $\mathsf{ID} \in \{0,1\}^*$ as its seed.

The system params are $(q, \mathbb{G}_1, \mathbb{G}_2, e, \mathbb{PV}, H_0, H_1, RNG)$. The master key is $\mathbb{SV}$. The message space is $\mathcal{M} = \{0,1\}^n$.

KeyGen. For a given identity $\mathsf{ID} \in \{0,1\}^*$, the algorithm performs the following:

- Compute $H_0(\mathsf{ID}) \to \{s_1, \cdots, s_t\}$.
- Seeding RNG with ID, generate pseudorandom sequences $(\alpha_1, \cdots, \alpha_t, \beta_{11}, \cdots, \beta_{tt}) \in \mathbb{F}_q$.
- Compute the private key as

$$x_{\mathsf{ID}} = \sum_{1}^{t} \alpha_i d_{s_i} + \sum_{1 \le i \le j \le t} \beta_{ij} d_{s_i} d_{s_j}.$$

- Compute the corresponding public key as

$$y_{\mathsf{ID}} = H(\mathsf{ID}) = \prod_{i=1}^{t} e(U_{s_i}, P)^{\alpha_i} \prod_{1 \le i \le j \le t} e(U_{s_i}, U_{s_j})^{\beta_{ij}}.$$

Note that $x_{\mathsf{ID}} = \log_g(y_{\mathsf{ID}})$. This enables any discrete logarithm based public key system to be used as the encryption primitive [9].

Encrypt. A message $m \in \mathcal{M}$ is encrypted for ID as follows.

- Select $r \in \mathbb{Z}_q$ at random and compute $U = g^r$.
- Compute $K = H_1(y_{\mathsf{ID}}^r)$.
- Parse K as $K_1 || K_2$.
- Compute $V = \mathcal{E}_s(K_1, M)$, $W = \mathcal{T}(K_2, V)$.
- The ciphertext is (U, V, W).

(Note that y_{ID} can be computed in the same way as y_{ID} in the KeyGen algorithm is computed solely using a given identity and the public parameter. This is not mentioned in [9], so we clarify it in our description.)

Decrypt. To decrypt a ciphertext $C = (U, V, W)$, encrypted under ID, the secret key holder conducts the following.

- Compute $K_1 || K_2 = H_1(U^{x_{\mathsf{ID}}})$.
- Check whether $W \stackrel{?}{=} \mathcal{T}(K_2, V)$.
- If it passes the verification, then output $M = D_s(K_1, V)$. Else output reject.

4. ANALYSIS OF CHEN ET AL.'S IDENTITY-BASED ENCRYPTION

Theorem 1. *Chen et al.'s identity-based encryption scheme [9] is insecure in the sense of* IND-sID-CCA.

Proof. Let $\mathcal{A}$ be an adversary that is permitted to issue a polynomial number of extraction queries. Taking a closer look at the extraction queries, each query will provide the following knowledge to $\mathcal{A}$.

$$x_{\mathsf{ID}} = \sum_{1}^{t} \alpha_i d_{s_i} + \sum_{1 \le i \le j \le t} \beta_{ij} d_{s_i} d_{s_j}$$

for $H_0(\mathsf{ID}) \to \{s_1, \cdots, s_t\}$.

To avoid confusion, for each query k with ID_k, the set of pseudorandom sequences is denoted as $(\alpha_{1_k}, \cdots, \alpha_{t_k}, \beta_{11_k}, \cdots, \beta_{tt_k}) \in \mathbb{F}_q$. After executing a polynomial number of extraction queries, say m times[1], then $\mathcal{A}$ will acquire the following set of knowledge.

$$\begin{aligned}
x_{\mathsf{ID}_0} &= \sum_{1}^{t} \alpha_{i_0} d_{s_{i_0}} + \sum_{1 \le i \le j \le t} \beta_{ij_0} d_{s_{i_0}} d_{s_{j_0}} \\
x_{\mathsf{ID}_1} &= \sum_{1}^{t} \alpha_{i_1} d_{s_{i_1}} + \sum_{1 \le i \le j \le t} \beta_{ij_1} d_{s_{i_1}} d_{s_{j_1}} \\
&\vdots \quad \vdots \\
x_{\mathsf{ID}_m} &= \sum_{1}^{t} \alpha_{i_m} d_{s_{i_m}} + \sum_{1 \le i \le j \le t} \beta_{ij_m} d_{s_{i_m}} d_{s_{j_m}}
\end{aligned}$$

In the above equation, there are ℓ unknown variables (namely $d_1, \cdots, d_\ell$), since ℓ is the dimension of the secret vector $\mathbb{SV}$

[1]We note that m can be chosen to be very large.

selected in the SETUP algorithm. Hence, the above problem constitutes $\mathcal{MQ}(q, \ell, m)$ problem.

This problem is known to be NP-hard for any field. The complexity of MQ problem depends on the size of q. Nevertheless, we note the following.

When the number of equations m is the same as the number of unknowns, ℓ, then the best known algorithms are exhaustive search for small fields, and a Gröbner bases algorithm for large fields. Unfortunately, Gröbner bases algorithm has large exponential complexity, and in practice, it cannot solve a system with $\ell \geq 15$. Hence, to make the scheme secure, one might think to select $\ell \geq 15$ to avoid this attack. However, Kipnis and Shamir [13] introduced an algorithm called "relinearization". While the exact complexity of this algorithm is unknown, for sufficiently overdefined systems, it was expected to run in polynomial time. Furthermore, Courtois, Klimov, Patarin and Shamir created the XL algorithm [11], which is an improvement of the relinearization technique by Kipnis and Shamir. For all $0 < \epsilon < \frac{1}{2}$ and $m \geq \epsilon\ell^2$, XL and relinearization are expected to run in *polynomial* time of approximately $\ell^{\mathcal{O}(1/\sqrt{\epsilon})}$.

Using the adversary $\mathcal{A}$ which can recover the master key (the secret vector $\mathbb{SV}$), one can easily construct an adversary that can break Chen et al's scheme in the `IND-sID-CCA` sense. ∎

We do note that Chen et al. recommended a fairly large number for ℓ, say 512, hoping to prevent the attack presented above. However, what we argue in this paper is that unless exponentially large ℓ is used, the `IND-sID-CCA` of Chen et al's scheme can be broken by the *polynomial-time* adversary, which invalidates the security result (proof) presented in [9]. (In other words, the recommended size 512 for ℓ is not sufficient to prevent the attack but if ℓ is exponentially large, the scheme can not be used in practice.)

We also note that independently, in the recent work in [2], Albrecht and Patterson demonstrated a refined attack which is more efficient than ours.

We remark that the difficulty in adopting the structure of the MQ problem to the key extraction algorithm of the identity-based encryption stems from the adversary's ability to make a number of key extraction queries at will. Using the private keys he has queried, the adversary can form a set of of equations (in polynomial time) to apply the XL algorithm. In order to to prevent the attack, one can choose l which is exceptionally larger (almost exponentially larger) than m. However, in this case, the size of $\mathbb{PV}(= (d_1P, \cdots, d_\ell P))$ should also be exponentially large, which makes Chen et al.'s identity-based encryption scheme impractical (if not impossible to use).

Based on the above theorem, we obtain the following corollary.

COROLLARY 1. *In contrast to the claim made in [9], the extract algorithm in [9] implies an insecure short signature scheme.*

5. CONCLUSION

In ASIACCS 2010, Chen, Charlemagne, Guan, Hu and Chen [9] proposed an interesting identity-based encryption scheme based on DHIES. They claimed that their scheme is secure in the sense of `IND-sID-CCA`. In this paper, we demonstrated that unfortunately, their scheme (which is either in the original description or in our modification) is insecure. The implication of our analysis is that to fix the scheme, exponentially large public parameter should be used but this is impossible to be realized in practice.

Acknowledgement. The authors would like to thank the anonymous referees of ASIACCS 2011 for the suggestions to improve this paper. We would also like to thank Prof. Kenny Paterson for some insight discussion.

6. REFERENCES

[1] M. Abdalla, M. Bellare, and P. Rogaway. The Oracle Diffie-Hellman Assumptions and an Analysis of DHIES. In D. Naccache, editor, *CT-RSA*, volume 2020 of *Lecture Notes in Computer Science*, pages 143–158. Springer, 2001.

[2] M. R. Albrecht and K. G. Paterson. Breaking an identity-based encryption scheme based on dhies. Cryptology ePrint Archive, Report 2010/637, 2010. `http://eprint.iacr.org/`.

[3] M. Bellare and P. Rogaway. Minimizing the use of random oracles in authenticated encryption schemes. In Y. Han, T. Okamoto, and S. Qing, editors, *ICICS*, volume 1334 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 1997.

[4] D. J. Bernstein, J. Buchmann, and E. Dahmen. Post-quantum cryptography. *Springer*, 2009.

[5] D. Boneh and X. Boyen. Efficient selective-id secure identity-based encryption without random oracles. In C. Cachin and J. Camenisch, editors, *EUROCRYPT*, volume 3027 of *Lecture Notes in Computer Science*, pages 223–238. Springer, 2004.

[6] D. Boneh and X. Boyen. Secure identity based encryption without random oracles. In M. K. Franklin, editor, *CRYPTO*, volume 3152 of *Lecture Notes in Computer Science*, pages 443–459. Springer, 2004.

[7] D. Boneh and M. K. Franklin. Identity-based encryption from the weil pairing. *SIAM J. Comput.*, 32(3):586–615, 2003.

[8] X. Boyen. A tapestry of identity-based encryption: practical frameworks compared. *IJACT*, 1(1):3–21, 2008.

[9] Y. Chen, M. Charlemagne, Z. Guan, J. bin Hu, and Z. Chen. Identity-based encryption based on DHIES. In D. Feng, D. A. Basin, and P. Liu, editors, *ASIACCS*, pages 82–88. ACM, 2010.

[10] C. Cocks. An identity based encryption scheme based on quadratic residues. In B. Honary, editor, *IMA Int. Conf.*, volume 2260 of *Lecture Notes in Computer Science*, pages 360–363. Springer, 2001.

[11] N. Courtois, A. Klimov, J. Patarin, and A. Shamir. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In B. Preneel, editor, *EUROCRYPT*, volume 1807 of *Lecture Notes in Computer Science*, pages 392–407. Springer, 2000.

[12] C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In C. Dwork, editor, *STOC*, pages 197–206. ACM, 2008.

[13] A. Kipnis and A. Shamir. Cryptanalysis of the hfe public key cryptosystem by relinearization. In M. J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture*

Notes in Computer Science, pages 19–30. Springer, 1999.

[14] B. Libert and J.-J. Quisquater. Identity based encryption without redundancy. In J. Ioannidis, A. D. Keromytis, and M. Yung, editors, *ACNS*, volume 3531 of *Lecture Notes in Computer Science*, pages 285–300, 2005.

[15] C. Peikert. Some recent progress in lattice-based cryptography. In O. Reingold, editor, *TCC*, volume 5444 of *Lecture Notes in Computer Science*, page 72. Springer, 2009.

[16] R. Sakai, K. Ohgishi, and M. Kasahara. Cryptosystems based on pairings. *Symposium on Cryptography and Information Security (SCIS 2000)*, 2000.

[17] A. Shamir. Identity-based cryptosystems and signature schemes. In *CRYPTO*, pages 47–53, 1984.

[18] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *FOCS*, pages 124–134. IEEE, 1994.

[19] N. P. Smart. Access control using pairing based cryptography. In M. Joye, editor, *CT-RSA*, volume 2612 of *Lecture Notes in Computer Science*, pages 111–121. Springer, 2003.

[20] B. Waters. Efficient identity-based encryption without random oracles. In R. Cramer, editor, *EUROCRYPT*, volume 3494 of *Lecture Notes in Computer Science*, pages 114–127. Springer, 2005.

[21] C. Wolf and B. Preneel. Applications of multivariate quadratic public key systems. In H. Federrath, editor, *Sicherheit*, volume 62 of *LNI*, pages 413–424. GI, 2005.