Comments on "An Efficient Identity-Based Provable Data Possession Protocol With Compressed Cloud Storage"

Lidong Han, Guangwu Xu, *Senior Member, IEEE*, and Qi Xie

***Abstract*— This article addresses some security issues of an identity-based provable data possession protocol with compressed cloud storage (Yang et al., 2022). Some serious flaws are identified and an attack to the protocol is designed. This attack is able to recover the ephemeral secret keys from two encrypted blocks with high probability to reveal the original plaintext file completely. Moreover, an adversary can impersonate a data owner to outsource any file to the cloud in a malicious way. The main ingredients of the attack is some classical number theoretic results.**

***Index Terms*— Cryptanalysis, provable data possession, ephemeral secret key, data auditing.**

I. Introduction

Provable data possession introduced by Ateniese et al. [2] is a technique which allows users to check the integrity of data stored on an untrusted cloud. After that, many research works [3], [4] discussed how to lower computational complexity to improve security and dynamic operations for public auditing of outsourced data. On the other hand, several previous schemes in [5] and [6] focused on users' public key generation without the help of public key infrastructure by constructing identity-based PDP schemes to facilitate certificate management.

Recently, Yang et al. proposed an identity-based PDP scheme, called IBPDP-CCS, to support compressed cloud storage [7]. The design of their protocol only relies on the basic algebraic operations and the costs of storage, communication, and computation are lowered. Specially, a data owner only needs to upload the encrypted blocks and a tag to the cloud without including his original file. The security of IBPDP-CCS was analyzed in [7] with the claim that "Except for the data owner, even though all other entities are collusive, they still cannot obtain the original user blocks."

In this article, some serious security flaws of the identity-based provable data possession protocol IBPDP-CCS are identified. More specifically, we demonstrate that in IBPDP-CCS, an attacker is able to use a piece of public information in the file tag to derive an ephemeral secret value by using two encrypted blocks with high probability. With this, other ephemeral private key values can be obtained as well. In particular, the original plaintext file is revealed completely and impersonating a data owner is possible. The attack is based on some classical result in number theory.

Manuscript received 3 June 2022; revised 19 April 2023; accepted 24 April 2023. Date of publication 27 April 2023; date of current version 5 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U21A20466, Grant 61972124, and Grant 12271306; and in part by the National Key Research and Development Program of China under Grant 2018YFA0704702. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrew Clark. *(Corresponding authors: Lidong Han; Guangwu Xu.)*

Lidong Han and Qi Xie are with the Key Laboratory of Cryptography of Zhejiang Province, Hangzhou 311121, China, and also with Hangzhou Normal University, Hangzhou 311121, China (e-mail: ldhan@hznu.edu.cn; qixie68@126.com).

Guangwu Xu is with the School of Cyber Science and Technology, Shandong University, Qingdao 266237, China, and also with the Quancheng Laboratory, Jinan 250103, China (e-mail: gxu4sdq@sdu.edu.cn).

Digital Object Identifier 10.1109/TIFS.2023.3271272

II. A Review of IBPDP-CCS

In this section, we briefly review the underlying identity-based PDP protocol proposed by Yang et al. [7] which achieves compressed cloud storage and contains four entities: data owner, cloud, third-party auditor (TPA), and key generation center (KGC). The IBPDP-CCS scheme consists of seven algorithms: Setup, Extract, Outsource, Challenge, ProofGen, Verify, and Recover. For more details, the readers are referred to [7].

1) $Setup(\lambda) \rightarrow (MSK, PK)$. With the security parameter λ, KGC determines a large prime p, generates a random number q with q being much smaller than p, and two random elements $g, \sigma \in Z_p$. A hash function $H : \{0,1\}^* \rightarrow Z_p$ is fixed. The master secret key MSK of the KGC is σ, and the public key is $PK = \{p, q, g, g^\sigma, H\}$.
2) $Extract(ID) \rightarrow SK_{ID}$. In this algorithm, KGC outputs the secret key SK_{ID} for a user whose identity is ID and the user validate it.
   - From a user identity ID, KGC selects a random number $\zeta \in Z_p$ and compute $a' = \zeta + \sigma H(ID) \mod (p-1)$. KGC transmits $SK_{ID} = a'$ to the user, together with g^ζ.
   - After receiving a' and g^ζ, the user determines $g^{a'} = g^\zeta \cdot g^{\sigma H(ID)} \mod p$ to judge the correctness of his secret key.
3) $Outsource(F, SK_{ID}, PK) \rightarrow (T, \tau)$. The user encrypts all blocks of the file F and generates the file tag.
   - The data owner randomly chooses $a'' \in Z_p, b, c, r, l \in Z_q$ and computes $a = a' + a''$, and $\hat{a} = a/r$. Note $ql \ll \hat{a}$.
   - The user divides the file into $\{x_1, x_2, \cdots, x_m\}$ and generates the encrypted block y_i by computing $y_i = a(x_i + bH(name||i)) + cx_i$, where $x_i \in Z_l$, $name$ is the identifier of the file F.
   - Define $T = \{y_1, y_2, \cdots, y_m\}$ as a set of all encrypted blocks. The owner generates the file tag $\tau = name||l||m||\hat{a}||g^a||g^c||g^{abc}||spk||SSig(name||l||m||\hat{a}||g^a||g^c||g^{abc}, ssk)$, where $SSig$ is an identity-based secure digital signature whose public key and secret key are spk and ssk respectively.
4) $Challenge(\cdot) \rightarrow chal$. TPA produces a challenge $chal$ when he wants to perform data audit.
   - TPA first checks the validity of the file tag τ with public key of ID-based signature. If invalid, TPA terminates the audit; otherwise, TPA extracts the values $m, l, \hat{a}, g^a, g^c, g^{abc}$ from the tag τ.
   - TPA chooses the random indices of the challenged block $\{i_1, i_2, \cdots, i_n\}$ from $[1, \cdots, m]$ and random numbers $\{e_1, e_2, \cdots, e_n\}$ such that $\sum_{j=1}^{n} e_j ql < \hat{a}$.
   - TPA sends to cloud the challenge sequence as $chal = \{i_1, i_2, \cdots, i_n; e_1, e_2, \cdots, e_n\}$.
5) $ProofGen(T) \rightarrow \Gamma$. The cloud generates a proof by computing $\Gamma = \sum_{j=1}^{n} e_{i_j} y_{i_j}$ as response to TPA.

6) *Verify*$(chal, \Gamma, \tau) \rightarrow v$. On receiving Γ, TPA verifies the equation

$$g^{cfloor(\Gamma/\hat{a})\cdot\hat{a}} \stackrel{?}{=} g^{a(\Gamma - floor(\Gamma/\hat{a})\cdot\hat{a})} \cdot g^{abc\sum_{j=1}^{n} e_j H(i_j)} \bmod p$$

If yes, TPA returns $v = 1$; otherwise, $v = 0$.

7) *Recover*$(y_i) \rightarrow x_i$. From y_i, the user can recover the original data block x_i by calculating $x_i = (y_i - floor(y_i/\hat{a}) \cdot \hat{a})/c$.

Remark: We would like to make some remarks on the definition and process of IBPDP-CCS. First we note that in [7], the authors' usage of $\hat{a} = a/r$ seems to be that of $floor$ as they also used the division / to result the *mod* operation.[1] However, the operator $floor$ was defined and used in many places. The second remark is about the decryption formulas $x_i = (y_i - floor(y_i/\hat{a}) \cdot \hat{a})/c,\ i = 1, 2, \cdots, m$ in [7]. We believe that some conditions need to be specified in order to make these equalities hold. In our rest discussion, we shall not need to involve those conditions. The decryption formulas are sufficient for us to perform analysis on IBPDP-CCS.

III. Security Analysis of IBPDP-CCS

In this section, we present an analysis of the IBPDP-CCS scheme. With achieving compressed cloud storage in mind, a data owner only transmits encrypted values to cloud without the original file to support integrity auditing and decryption. However, we are able to describe an attack to IBPDP-CCS in which an adversary can decrypt all encrypted blocks of files. Specifically, an attacker is able to recover the ephemeral private key the data owner used to perform decryption. The main ingredients of the attack are some basic number theoretic primitives such as computing the greatest common divisor (GCD) and a classical result of Dirichlet [8] which states.

Theorem 1: If α and β are two random integers, the probability that $gcd(\alpha, \beta) = 1$ is $\frac{6}{\pi^2} \approx 0.608$.

As usual, an adversary is assumed to have the ability of eavesdropping the information from a communication channel or infected servers. This has been specified by the security model in [7], which allows the adversary to query the values of user's partial key and encrypted blocks and other information like file tags. A further protection by using standard protocols for confidentiality and integrity as in the security link assumption of [9] might help to mitigate the attack, but there are issues of efficiency and integrating of protocols.

Now let us describe the attack in details. With the set of encrypted blocks $T = \{y_1, y_2, \cdots, y_m\}$ and a file tag τ, the adversary performs the following steps.

- *Step* 1. The adversary selects two random blocks y_i, y_j, and extracts the value $\hat{a}$ from the file tag τ.
- *Step* 2. From y_i, y_j and $\hat{a}$, the adversary computes
  $\alpha_i = y_i - floor(y_i/\hat{a}) \cdot \hat{a}$
  $\alpha_j = y_j - floor(y_j/\hat{a}) \cdot \hat{a}$
- *Step* 3. The adversary calculates $c' = \gcd(\alpha_i, \alpha_j)$ using the Euclidean algorithm and then calculates $x'_k = \big(y_k - floor(y_k/\hat{a}) \cdot \hat{a}\big)/c'$ for $k = 1, 2, \cdots, m$.
- *Step* 4. The adversary checks whether the file $\{x'_1, x'_2, \cdots, x'_m\}$ is meaningful. If yes, then the private key c' is valid. Otherwise, go to Step 1.

A proof of the correctness of the above procedure goes as follows. According to the design and requirements of the IBPDP-CCS protocol, the equality $y_i = floor(y_i/\hat{a}) \cdot \hat{a} + cx_i$ holds for each $1 \le i \le m$. So

$$\gcd(\alpha_i, \alpha_j) = c \gcd(x_i, x_j).$$

It can be assumed that the integers x_i (converted from the actual file blocks) exhibit some randomness. Thus with bigger probability $\gcd(x_i, x_j) = 1$ holds true by Theorem 1.

Remark: In fact, one can do better than the described protocol. Computing $c'_{i,j} = \gcd(\alpha_i, \alpha_j)$ for all $1 \le i, j \le m, i \ne j$, then with overwhelming probability $c = \min_{1\le i,j\le m, i\ne j} c'_{i,j}$. The worst situation is that all $\gcd(x_i, x_j) > 1$, then $\min_{1\le i,j\le m,i\ne j} c'_{i,j}$ is still likely a small multiple of c.

We would like to further remark that, once an adversary obtains the valid value c, he can recover the other private key a, b using the similar technique as above. More specifically, from the equation $y_i = a(x_i + bH(name||i)) + cx_i$, an adversary has known the values of y_i and $cx_i = y_i - floor(y_i/\hat{a})$. Then, for a pair (y_i, y_j), he generates $\gcd(y_i - cx_i, y_j - cx_j)$ which is probably the values of a from the aforementioned analysis. Then b is recovered using $a, c, x_i, y_i, H(name||i)$. Therefore, an adversary can impersonate the owner to outsource any file which has same file tag τ in a malicious manner.

IV. Conclusion

This article presents an analysis of the PDP protocol IBPDP-CCS [7]. Some serious security flaws are identified and an attack to IBPDP-CCS is described. An attacker is able to recover all encrypted blocks with high probability without knowing the secret key of the owner. Furthermore, an adversary can impersonate the data owner to outsource files to the cloud.

[1] If '/' were for rational division, the secret a would be easily derived from the public piece $\hat{a}$.

References

[1] Y. Yang, Y. Chen, F. Chen, and J. Chen, "An efficient identity-based provable data possession protocol with compressed cloud storage," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1359–1371, 2022, doi: 10.1109/TIFS.2022.3159152.

[2] G. Ateniese et al., "Provable data possession at untrusted stores," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, Oct. 2007, pp. 598–609.

[3] C. Erway, A. Küpçü, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proc. 16th ACM Conf. Comput. Commun. Secur.*, Nov. 2009, pp. 213–222.

[4] Q. Wang et al., "Enabling public verifiability and data dynamics for storage security in cloud computing," in *Proc. ESORICS*, 2009, pp. 355–370.

[5] H. Wang, Q. Wu, B. Qin, and J. Domingo-Ferrer, "Identity-based remote data possession checking in public clouds," *IET Inf. Secur.*, vol. 8, no. 2, pp. 114–121, Mar. 2014.

[6] J. Li, H. Yan, and Y. Zhang, "Certificateless public integrity checking of group shared data on cloud storage," *IEEE Trans. Services Comput.*, vol. 14, no. 1, pp. 71–81, Jan. 2021.

[7] Y. Yang et al., "An efficient identity-based provable data possession protocol with compressed cloud storage," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1359–1371, 2022.

[8] G. L. Dirichlet, *Über die Bestimmung der Mittleren Werthe der Zahlentheorie*. Berlin, Germany: Abhandlungen der Köiglichen Akademie der Wissenchaften zu Berlin, 1849.

[9] N. D. Sarier, "Improving the accuracy and storage cost in biometric remote authentication schemes," *J. Netw. Comput. Appl.*, vol. 33, no. 3, pp. 268–274, May 2010.