A Cryptanalysis of the Original Domingo-Ferrer's Algebraic Privacy Homomorphism

Jung Hee Cheon and Hyun Soo Nam

School of Mathematical Sciences, Seoul National University, Republic of Korea {jhcheon, hsnam}@math.snu.ac.kr

Abstract. We propose a cryptanalysis of the original Domingo-Ferrer's algebraic privacy homomorphism. We show that the scheme over \mathbb{Z}_n can be broken by d+1 known plaintexts in $O(d^3 \log^2 n)$ time when it has d times expansion through the encryption. Furthermore even when the public modulus n is kept secret, it can be broken by d+2 known plaintexts in time at most $O(d^5 \log^2(dn))$.

Keywords: Privacy homomorphism, Encrypted Data, Database Security

1 Introduction

Recently, rapid advances in networking and internet have introduced Application Service Provider (ASP) as a new e-business model. Software as a service includes rent-a-spreadsheet, electronic mail services, general storage services, disaster protection services, database as a service, etc. In the ASP model including a database service model, there are two main privacy issues. First, the owner of the data needs to be assured that the data stored on the service-provider site is protected against data modification or eavesdropping from unauthorized outsiders. Second, data need to be protected even from the service providers, if the providers themselves cannot be completely trusted. The first requirement can be achieved by access control. But the second one is not easy to be satisfied by the similar technique.

The concept of processing encrypted data was firstly introduced by Rivest, Adleman, and Dertouzos [9] in 1978 in order to resolve the second problem. They proposed several privacy homomorphisms to process encrypted data without decrypting. A privacy homomorphism is an encryption function which allows processing the

1 INTRODUCTION 2

encrypted data without knowledge of the decryption function. Further, an algebraic privacy homomorphic encryption \mathcal{E} over a ring \mathcal{R} is an encryption function which has efficient algorithms to compute $\mathcal{E}(xy)$ and $\mathcal{E}(x+y)$ from $\mathcal{E}(x)$ and $\mathcal{E}(y)$ without revealing x and y. One example of an algebraic privacy homomorphism [9] is as follows:

Example 1. Let p, q be large primes, and n = pq.

$$\mathcal{P} = \mathbb{Z}_n, \quad \mathcal{C} = \mathbb{Z}_p \times \mathbb{Z}_q, \quad \mathcal{E}(x) = (x \mod p, x \mod q)$$

and decryption is done using the Chinese remainder theorem.

Apparently, this function is an algebraic privacy homomorphism under the usual modular addition and modular multiplication. Unfortunately, it is shown in [4] that this algebraic privacy homomorphism can be broken using a known-plaintext attack.

In 1991, Feigenbaum and Merritt [8] questioned directly whether an algebraic privacy homomorphism does exist. In spite of numerous studies over twenty five years, little progress has been made in deciding whether or not an algebraic privacy homomorphism exists. Neither promising candidates for such schemes nor evidence that such schemes exist has been found. Many suggested examples and schemes are shown to be insecure [4, 2, 12]. Ahituv et al. showed that any algebraic privacy homomorphism can be broken efficiently by chosen ciphertext attacks [1]. Boneh and Lipton proved that any deterministic algebraic privacy homomorphism over rings \mathbb{Z}_n can be broken in sub-exponential time under a (reasonable) number theoretic assumption [3].

Domingo-Ferrer proposed two algebraic privacy homomorphisms in 1996 and 2002 [6, 7]. The second one is broken by Wagner and Bao [12, 2]. But there is no serious attack on the first scheme. It is the only algebraic privacy homomorphism that remains secure to the authors' best knowledge. In this paper, we analyze the original privacy homomorphism of Domingo-Ferrer [6], and show that it is not secure against known-plaintext attacks. More precisely, when we consider an encryption function from \mathbb{Z}_n to $(\mathbb{Z}_p \times \mathbb{Z}_q)^d$ for n = pq, it can be broken by d+1 plaintext-ciphertext pairs in $O(d^3 \log^2 n)$. Even when n is kept secret, it can be broken by one more pairs in $O(d^5 \log(dn))$ with almost 1 probability. It means that one can increase the security by increasing d. However, the efficiency decreases as d increases. Moreover, if we want a scheme which is secure against at most d ciphertext, linear transformation are better in efficiency aspects.

The outline of the paper is as follows: In Section 2, we introduce the original scheme. In Section 3, we propose its attack using d+1 known plaintexts. In Section 4, we propose the attack of the Domingo-Ferrer scheme when the public modulus is kept secret. This condition was proposed by the author to increase the security and

reduce the efficiency. But we show that this enhanced version also can be broken by one more known plaintext than the original version with very high probability. We conclude in Section 5.

2 Domingo-Ferrer algebraic privacy homomorphism

Let me introduce the original Domingo-Ferrer scheme [6].

The Domingo-Ferrer scheme

Let p, q be large primes with p < q, and n = pq. For a positive integer d, we set

$$\mathcal{P} = \mathbb{Z}_n, \quad \mathcal{C} = (\mathbb{Z}_p \times \mathbb{Z}_q)^d.$$

- 1. **Public Parameter** d and n. n can be kept secret to increase the security at the sacrifice of the efficiency.
- 2. Secret key p, q and randomly chosen integer $r_p \in \mathbb{Z}_p^*$ and $r_q \in \mathbb{Z}_q^*$, each of which generates a large multiplicative subgroup of \mathbb{Z}_p^* and \mathbb{Z}_q^* , respectively.
- 3. **Encryption** Randomly split $x \in \mathbb{Z}_n$ into secret $x_1, x_2, \dots x_d$ such that

$$x = \sum_{i=1}^{d} x_i \mod n$$
 and $x_i \in \mathbb{Z}_n$.

$$\mathcal{E}_k(x) = ([x_1 r_p \bmod p, x_1 r_q \bmod q], [x_2 r_p^2 \bmod p, x_2 r_q^2 \bmod q], \dots, [x_d r_p^d \bmod p, x_d r_q^d \bmod q])$$

4. **Decryption** Compute the scalar product of the j-th [mod p, mod q] pair by $[r_p^{-j} \text{mod } p, r_q^{-j} \text{mod } q]$ to retrieve the $[x_j \text{mod } p, x_j \text{mod } q]$. Add them up to get [x mod p, x mod q]. Use the Chinese remainder theorem to get x mod n.

Domingo-Ferrer's scheme is an algebraic privacy homomorphism with respect to the operation of addition and multiplication in \mathbb{Z}_n and encrypted values are computed over $(\mathbb{Z}_n \times \mathbb{Z}_n)^d$. Note that if d = 1, $r_p = 1$, and $r_q = 1$, then Domingo-Ferrer's algebraic privacy homomorphism scheme is the same with Example 1.

Domingo-Ferrer claims that his scheme has the following improvements from Example 1.

- Small values are nontrivially encrypted while the small value plaintext is the same with ciphertext component in Example 1.
- It is able to withstand a known-plaintext attack.

We will show that the second assertion is false.

3 Security analysis of Domingo-Ferrer's system

Let $x \in \mathbb{Z}_n$ be a plaintext such that

$$x \equiv x_1 + x_2 + \dots x_d \operatorname{mod} n \tag{1}$$

and

$$\mathcal{E}(x) = ([x_1 r_p \mod p, x_1 r_q \mod q], \dots, [x_d r_p{}^d \mod p, x_d r_q{}^d \mod q])$$

= $([y_1, z_1], [y_2, z_2], \dots, [y_d, z_d]).$

We denote $\mathcal{E}_p(x) = (y_1, \dots, y_d)$ and $\mathcal{E}_q(x) = (z_1, \dots, z_d)$. For a subset S of \mathbb{Z}_n , if $\{(x, \mathcal{E}_p(x)) | x \in S\}$ and $\{(x, \mathcal{E}_q(x)) | x \in S\}$ are linearly independent as a module elements of \mathbb{Z}_n^{d+1} over \mathbb{Z}_n then we say that $\{(x, \mathcal{E}(x) | x \in S\}$ is linearly independent over \mathbb{Z}_n .

LEMMA 2. Let $k \leq d+1$. Given randomly chosen k plaintext-ciphertext pairs (M_i, C_i) for $i = 1, \ldots, k$, the probability that the set $\{(M_i, C_i) | i = 1, \ldots, k\}$ is linearly independent is more than $e^{-\frac{4}{p-1}}$.

Proof. Note that

$$Prob\{(M_1, C_1), \dots, (M_k, C_k) \text{ are linearly independent over } \mathbb{Z}_p\}$$

$$= \frac{(p^{d+1} - 1)(p^{d+1} - p)(p^{d+1} - p^2) \cdots (p^{d+1} - p^{k-1})}{(p^{d+1})^k}$$

$$= (1 - \frac{1}{p^{d+1}})(1 - \frac{1}{p^d}) \cdots (1 - \frac{1}{p^{d-k+1}})$$

$$\geq (1 - \frac{1}{p^{d+1}})(1 - \frac{1}{p^d}) \cdots (1 - \frac{1}{p^{d-k+1}}) \cdots (1 - \frac{1}{p})$$

$$\geq e^{-2(\frac{1}{p^{d+1}} + \frac{1}{p^d} + \cdots \frac{1}{p})} \quad (\because 1 - x \ge e^{-2x}, 0 \le x \le \frac{1}{2})$$

$$\geq e^{-\frac{2}{p-1}}.$$

When we assume p < q,

$$Prob\{(M_1, C_1), \dots, (M_k, C_k) \text{ are linearly independent over } \mathbb{Z}_n\}$$

$$= Prob\{(M_1, C_1), \dots, (M_k, C_k) \text{ are linearly independent over } \mathbb{Z}_p \text{ and } \mathbb{Z}_q\}$$

$$\geq e^{-\frac{2}{p-1}} e^{-\frac{2}{q-1}} \geq e^{-\frac{4}{p-1}}.$$

Note that if $(M_1, C_1), \ldots, (M_k, C_k)$ are linearly dependent over \mathbb{Z}_n , one of the pairs, for example (M_i, C_i) , can be computed by a linear combination of the others. That means that if we have all the other pairs, we can decrypt C_i to get M_i . The converse is true in the sense that one pair can be used to generate many linearly dependent pairs. Thus it is reasonable to discard linearly dependent pairs over \mathbb{Z}_n .

Preparation

Assume we have a plaintext-ciphertext pair $(x, ([y_1, z_1], \dots, [y_d, z_d]))$ where $x \equiv x_1 + x_2 + \dots + x_d \mod n$, $y_i \equiv r_p^i x_i \mod p$ and $z_i \equiv r_q^i x_i \mod q$. Since p divides n, we have

$$x \equiv x_1 + x_2 + \dots + x_d \operatorname{mod} p. \tag{2}$$

By replacing $y_i \equiv r_p^i x_i \mod p$ in (2), we obtain

$$-x + y_1 t + y_2 t^2 + \dots + y_d t^d \equiv 0 \operatorname{mod} p$$
(3)

where $t = r_p^{-1} \mod p$.

Using the equation (3), we will show that this system can be broken by a known-plaintext attack. In Section 3.1, we deal with the case that the modulus n is public, and in Section 3.2 we deal with the case that the modulus n is kept secret.

3.1 The case that n is public

Assume that a cryptanalyst has d+1 linearly independent plaintext-ciphertext pairs (M_i, C_i) , i = 1, 2, ..., d+1 over \mathbb{Z}_n . Let $C_i = ([y_{i1}, z_{i1}], [y_{i2}, z_{i2}], ..., [y_{id}, z_{id}])$.

Step 1 Finding p, q

By applying the equation (3) to d+1 pairs $(M_1, C_1), \ldots, (M_{d+1}, C_{d+1})$, we get the following d+1 modular equations:

$$\begin{array}{rcl}
-M_{1} + y_{11}t + y_{12}t^{2} + \dots + y_{1d}t^{d} & \equiv 0 \mod p, \\
-M_{2} + y_{21}t + y_{22}t^{2} + \dots + y_{2d}t^{d} & \equiv 0 \mod p, \\
& \vdots \\
-M_{d+1} + y_{d+11}t + y_{d+12}t^{2} + \dots + y_{d+1d}t^{d} & \equiv 0 \mod p.
\end{array} \tag{4}$$

This can be transformed into the following matrix form:

$$\begin{pmatrix} M_1 & y_{11} & y_{12} & \cdots & y_{1d} \\ M_2 & y_{21} & y_{22} & \cdots & y_{2d} \\ & \vdots & & & \\ M_{d+1} & y_{d+11} & y_{d+12} & \cdots & y_{d+1d} \end{pmatrix} \begin{pmatrix} -1 \\ t \\ \vdots \\ t^d \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \operatorname{mod} p. \tag{5}$$

Since the homogeneous equation (5) has a nontrivial solution in \mathbb{Z}_p^d , the coefficient matrix

$$A = (M|C) = \begin{pmatrix} M_1 & y_{11} & y_{12} & \cdots & y_{1d} \\ M_2 & y_{21} & y_{22} & \cdots & y_{2d} \\ & \vdots & & & \\ M_{d+1} & y_{d+11} & y_{d+12} & \cdots & y_{d+1d} \end{pmatrix}$$

has $\det(A) = 0 \mod p$. If d+1 pairs (M_i, C_i) are linearly independent, we have $\det(A) \neq 0 \mod n$. Therefore $\gcd(\det(A) \mod n, n) = p$. q is obtained from q = n/p.

Step 2 Finding r_p, r_q

Once p, q are known, r_p, r_q can be computed by solving the system of linear equations (5) over \mathbb{Z}_p . Hence the solution can be found in $O(d^3 \log^2 p)$ using Gaussian elimination.

¿From Step 1 and Step 2, we can compute all secret keys using d+1 linearly independent plaintext-ciphertext pairs. Further the probability that randomly chosen k pairs are linearly dependent is very large. More precisely, we obtain the following theorem.

Theorem 3. Domingo-Ferrer's algebraic privacy homomorphism with public modulus can be broken by linearly independent d+1 known plaintext-ciphertext pairs in time $O(d^3 \log^2 n)$. Further randomly chosen d+1 plaintext-ciphertext pairs are linearly independent with the probability larger than $e^{\frac{-4}{p-1}}$.

Remark that r_p can be computed from the equation (4). Finding a root for a polynomial of degree d in \mathbb{Z}_p takes $O(d^2 \log d \log^3 p)$ using Berkelamp-Rabin algorithm. If d is larger than $\log p$, we may use this algorithm rather than Step 2.

3.2 The case that n is not public

Domingo-Ferrer suggested that his scheme could be a more secure scheme by hiding the modulus n. In this case, however, we can not perform many multiplications since one multiplication doubles up the length of ciphertexts. We will show that even if n is not public, a similar cryptanalysis can be applied.

Let me assume we have d+2 known plaintext-ciphertext pairs which are linearly independent over \mathbb{Z} , not in \mathbb{Z}_n . From each d+1 pairs, we can induce the $(d+1) \times (d+1)$ coefficient matrix A_i for $i=1,2,\ldots,d+2$ as in the equation (5).

Computing the determinant of an integer matrix takes an exponential time using an ordinary Gaussian algorithm. Thus we need a special method as follows:

- 1. Estimate the bound R by $R = max\{y_{ij}, z_{ij}\}$.
- 2. det A_i is bounded by $M = (d+1)!R^{d+1}$ from the determinant formula.
- 3. If we compute $\det(A_i \mod M)$, it is the same with $\det A_i$ since it is smaller than M.

Then $gcd(det(A_1), det(A_2), det(A_3), \dots, det(A_k))$ where $2 \le k \le d+2$ will be p with high probability because the probability of

$$\gcd(\det(A_1)/p, \det(A_2)/p, \det(A_3)/p, \dots, \det(A_k)/p) = 1$$

can be estimated approximately $\frac{1}{\zeta(k)}$. More precisely, we have

LEMMA 4. Assume that divisibility of an integer by different primes is independent. Let N be a positive integer. If positive integers n_1, \ldots, n_k are randomly drawn from the interval (0, N), then the probability $P_k(d)$ of the greatest common divisor of k integers n_1, \ldots, n_k is equal d,

$$P_k(d) = \lim_{N \to \infty} Pr\{gcd(n_1, n_2, \dots n_k) = d\} \approx \frac{1}{d^k \zeta(k)}$$

where $\zeta(k)$ is Riemann's zeta function.¹

If $\ell = \gcd(\det(A_1), \det(A_2), \det(A_3), \dots, \det(A_k))$ is not prime and ℓ has only small factors other than p, then we can easily calculate p by trial divisions or some integer factoring algorithms. Using Lemma 4, we obtain that the probability for finding find p is

$$\sum_{d=1}^{m} P_k(d) = \sum_{d=1}^{m} \frac{1}{d^k \zeta(k)} = \frac{1}{\zeta(k)} \left(\sum_{d=1}^{m} \frac{1}{d^k}\right) = 1 - \frac{1}{\zeta(k)} \left(\sum_{d=m}^{\infty} \frac{1}{d^k}\right)$$

for $m=\ell/p$. This value approaches 1 as m increases and is enough close to 1 even for $m=2^{30}$.

We can find q by the same method. After getting p, q, we can compute the other secret keys r_p, r_q by the same method with the case that n is known.

.6079271016,.9239384016,.9829525910,.9959391987,.9990064106

by using the Euler's formula

$$2\zeta(2\eta) = (-1)^{m+1} \frac{(2\pi)^{2\eta}}{(2\eta)!} B_{2\eta}$$

where $B_{2\eta}$ is the Bernoulli numbers.

¹The approximate value of $\frac{1}{\zeta(k)}$ for k=2,4,6,8,10 is

Theorem 5. Domingo-Ferrer's algebraic privacy homomorphism with secret modulus can be broken by d+2 known plaintext-ciphertext pairs which have two linearly independent subsets of d+1 plaintext-ciphertext pairs in time $O(d^5\log^2(dn)+\varepsilon)$ where ε depends on a suitable factorization algorithm and the probability of success is approximately $1(\approx \frac{1}{\zeta(2)}(\sum_{d=m}^{\infty} \frac{1}{d^2}))$. Further the probability that random d+2 plaintext-ciphertext pairs have two linearly independent d+1 plaintext-ciphertext pairs is approximately larger than $(e^{\frac{-4}{p-1}})^2$.

Proof. Assume we use only two matrices to save time for computing determinants in \mathbb{Z} . The computing time of $\det(A)$ in \mathbb{Z} is estimated as follows: Take a prime $M \simeq (d+1)! n^{d+1} \leq ((d+1)n)^{d+1}$ so that $|\det(A)| \leq M$. Compute $\det(A \mod M)$ using Gaussian elimination. So the complexity is $O((d+1)^3 \log^2 M)$ bit operations, which is approximately $O((d+1)^5 \log^2 ((d+1)n))$. Next, using a suitable factorization algorithm (whose complexity is ε), we can compute p. The complexity of computing r_p and r_q is the same with the known n case and so $O(d^3 \log^2 n)$. The total complexity is $O(d^5 \log^2(dn) + \varepsilon)$ And the probability that random d+2 plaintext-ciphertext pairs have two linearly independent d+1 plaintext-ciphertext pairs approximately larger $(e^{\frac{-4}{p-1}})^2$, and the probability of finding p from the $\gcd(A_1, A_2)$ is approximately 1 by the above explanation $1 - \frac{1}{\zeta(2)}(\sum_{d=m}^{\infty} \frac{1}{d^2}) \approx 1$ if $m \gg 2^{100}$

Case	n is known	n is unknown
# of known plaintexts	d+1	d+2
The condition of	linearly	two subsets with $d+1$ elements
plaintext-ciphertext pairs	independent	are linearly independent
The probability	$\geq e^{\frac{-4}{p-1}}$	$\gtrsim e^{\frac{-8}{p-1}}$
of the above condition		
Time	$O(d^3 \log^2 n)$	$O(d^5 \log^2(dn) + \varepsilon)$
Probability of success	1	≈ 1

Table 1: The Complexity of the Proposed Attacks

3.3 Baby example

This example is in the paper of Domingo-Ferrer [6]. We illustrate the above analysis by this example. Let p = 17, q = 13, $r_p = 2$, and $r_q = 3$ be secret key. We can

encrypt -1, 3, 1, and 2 as follows:

$$\mathcal{E}_k(-1) = \mathcal{E}_k(2, -3) = ([4, 6], [5, 12])$$

$$\mathcal{E}_k(3) = \mathcal{E}_k(2, 1) = ([4, 6], [4, 9])$$

$$\mathcal{E}_k(1) = \mathcal{E}_k(4, -3) = ([8, 12], [5, 12])$$

$$\mathcal{E}_k(2) = \mathcal{E}_k(3, -1) = ([6, 9], [13, 4]).$$

We use three plaintext-ciphertext pairs, for example plaintext (-1,3,1) then the matrix

$$A = \begin{pmatrix} -1 & 4 & 5 \\ 3 & 4 & 4 \\ 1 & 8 & 5 \end{pmatrix}, \ \det(A) = 68, \ \gcd(\det(A), 13 * 14 = 221) = 17.$$

If n = 221 is unknown, we use above A and use one more plaintext-ciphertext pair and make a matrix for plaintext (-1,3,2),

$$B = \begin{pmatrix} -1 & 4 & 5 \\ 3 & 4 & 4 \\ 2 & 6 & 13 \end{pmatrix}, \det(B) = -102, \gcd(\det(A), \det(B)) = 34$$

We next calculate r_p by using $1 + 4t + 5t^2 = 0, -3 + 4t + 4t^2 = 0$ then by elimination $-19 + 4t = 0 \mod 17$ so $t = 9 \mod 17, r_p = t^{-1} \mod 17 = 2$.

4 Conclusion and Open problems

We conclude this section by summarize Domingo-Ferrer's algebraic privacy homomorphism.

- 1. Addition, substraction, multiplication and division can be carried out on encrypted data by simple integer operations without decrypting.
- 2. A given plaintext can be encrypted into many ciphertext versions by using a padding function from \mathbb{Z}_n to \mathbb{Z}_n^d .
- 3. A ciphertext is about d times longer than the corresponding plaintext. If d is large then the storage leakage is inevitable.
- 4. It can be efficiently broken by d+1 plaintext-ciphertext pairs.

If we have at most d pairs, Domingo-Ferrer's scheme seems to be secure. But in this case, we might have more efficient scheme using linear transformations rather than the nonlinear transformation.

REFERENCES 10

References

[1] N. Ahituv, y. Lapid and S. Neumann, "Processing Encrypted Data," Communications of the ACM. Vol. 20, pp. 777-780, 1987.

- [2] F. Bao, "Cryptanalysis of a Provable Secure Additive and Multiplicative Privacy Homomorphism," To appear in ICSD2003, 2003.
- [3] D. Boneh and R. Lipton, "Searching for Elements in Black-Box Fields and Applications," In Advances in Cryptology-Crypto'96, LNCS1109, pp. 283-297, Springer-Verlag, 1996.
- [4] E. Brickell and Y. Yacobi, "On Privacy Homoorphisms," In Advances in Cryptology-Eurocrypt'87, pp. 117-125, Springer-Verlag, 1988.
- [5] H. Cohen, A Course in Computational Algebraic Number Theory, Springer-Verlag, Graduate Texts for Mathematics, Vol. 138, 1993
- [6] J. Domingo-Ferrer, "A New Privacy Homomorphism and Applications," Information Processing Letters, Vol. 60, no. 5, pp. 277-282, Dec.1996.
- [7] J. Domingo-Ferrer, "A Provably Secure Additive and Multiplicative Privacy Homomorphism," ISC2002, LNCS. Vol. 2443, pp.471-483, 2002.
- [8] J. Feigenbaum, M.Merritt, "Open Questions, Talk Abstracts, and Summanry of Discussions," DIMACS series in Discrete Mathematics and Theoretical Computer Science, Vol. 2, pp. 1-45, 1991.
- [9] R.L. Rivest, L. Adleman and M.L. Dertouzos, "On Data Banks and Privacy Homorphisms," In Foundataions of Secure Computation, pp. 169-179, Academic Press, 1978.
- [10] R.L. Rivest, A. Shamir and L. Adleman, "A Method for Obtaning Digital Signatures and Public Key Cryptosystems," Communications of ACM, Vol. 21, pp. 120-126, April 1978.
- [11] M. R. Schroeder, Number Theory in Science and Communication, 2nd ed. Berlin, Springer-Verlag, 1986.
- [12] D. Wagner. "Cryptanalysis of an Algebraic Privacy Homomophism," To appear in ISC2003, 2003.