On Privacy Homomorphisms
(Extended Abstract)

Ernest F. Brickell
Yacov Yacobi

Bell Communications Research
435 South Street
Morristown, New Jersey 07960

Abstract

An additive privacy homomorphism is an encryption function in which the decryption of a sum (or possibly some other operation) of ciphers is the sum of the corresponding messages. Rivest, Adleman, and Dertouzos have proposed four different additive privacy homomorphisms. In this paper, we show that two of them are insecure under a ciphertext only attack and the other two can be broken by a known plaintext attack. We also introduce the notion of an R-additive privacy homomorphism, which is essentially an additive privacy homomorphism in which only at most R messages need to be added together. We give an example of an R-additive privacy homomorphism that appears to be secure against a ciphertext only attack.

1. Introduction

A privacy homomorphism is an encryption function which allows the encrypted data to be operated on without knowledge of the decryption function. Privacy homomorphisms were introduced by Rivest, Adleman, and Dertouzos [RAD]. Secure privacy homorphisms could be applied to protect data-bases against eavesdropping by the system manager.

[RAD] mentioned that privacy homomorphisms could be used to establish data base systems in which an encrypted total of a list of items could be computed using only the encrypted values of the list of items. Privacy homomorphisms could also be used to establish a secure conference telephone call. In a typical conference call, a central facility adds together the signals of the speakers. If the signals were encrypted using an additive privacy homomorphism, then the central facility could "add" the signals together without decrypting them.

We must now give a more formal definition. A privacy homomorphism is a family of functions $(e_k, d_k, \alpha, \gamma)$ such that $d_k(\gamma(e_k(m_1), ..., e_k(m_r))) = \alpha(m_1, ..., m_r)$ for each k in key space **K** and any $m_1, ..., m_r$ in message space **M**. The

D. Chaum and W.L. Price (Eds.): Advances in Cryptology - EUROCRYPT '87, LNCS 304, pp. 117-125, 1988.
© Springer-Verlag Berlin Heidelberg 1988

definition given in [RAD] is more general. In considering the security of these systems, we will assume that the cryptanalyst knows the functions e_k, d_k, α, γ but does not know the key k. Ahituv, Lapid & Neumann [ALN] showed that any privacy homomorphism, which has addition as its ciphertext domain operation is insecure under chosen ciphertext attack.

One privacy homomorphism mentioned in [RAD] is based on the multiplicative property of the RSA encryption function [RSA]. Let $n = pq$ where p and q are large primes. Let $\alpha(m_1, \ldots, m_r) = m_1 \cdots m_r \bmod n$ and $\gamma = \alpha$. Define e and d in the usual manner for RSA. This privacy homomorphism is as secure as RSA.

There are four more privacy homomorphisms mentioned in [RAD]. There are weaknesses in each of these which will be discussed in section 2.

We now introduce the notion of an R-additive privacy homomorphism, in which only the addition of at most R messages is allowed and α is addition over the integers. Let M be a subset of the integers. An R-additive privacy homomorphism is a family of functions $(e_k, d_k, \alpha, \gamma)$ such that $d_k(\gamma(e_k(m_1), \ldots; e_k(m_r))) = m_1 + \cdots + m_r$ for $r \leq R$ and for each k in key space K and any $m_1, \ldots, m_r$ in message space M. We do not know of any example of an R-additive privacy homomorphism that it thought to be secure against a known plaintext attack. It would be very interesting to know whether such a function exists that is provably secure in the sense that cryptanalysis can be shown equivalent to some well studied problem.

An R-additive privacy homomorphism would provide a nice solution to the problem of designing a secure conference call [BLY]. For this application, it would only be necessary for R to be 2 or 3, but it would be desirable to limit the data expansion. For large enough R, R-additive privacy homomorphisms for which γ is addition will be insecure under the chosen ciphertext attack of [ALN]. However for small R, this attack does not apply. This will be explained in more detail in section 3. It is possible that a privacy homomorphism that is succeptible to a known plaintext attack or a chosen ciphertext attack could still be secure for encrypting digitized speech. Therefore, we think that it is worthwhile to examine the security of R-additive privacy homomorphisms against a ciphertext only attack.

In section 3, we will describe an example of an R-additive privacy homomorphism. At first glance, it appears that the attacks that broke the knapsack type cryptosystems would also break this system using a ciphertext only attack. In section 4, we will show that this is probably not the case. In section 5, we will describe how this system could be modified to possibly improve the security.

2. Cryptanalysis of [RAD] privacy homomorphisms

We present cryptanalysis of four privacy homomorphism systems which appear in [RAD] (examples 1,5,3,4 in this order).

System 1

Let g be a generator modulo a prime p, where $p-1=\prod_{i=1}^{k} p_i^{d_i}$, and for all $i, p_i \leq B$, for some small B. Let q be a large prime and let $n=p \cdot q$. Message M is encrypted by computing: $C \equiv g^M \bmod n$. Decryption is $M \equiv \log_g C$ $\bmod p$. The structure of p enables computation of the discrete logarithm by the method of Pohlig and Hellman [PH] in time $O(B^{1/2})$. This is a privacy homomorphism with α being addition $\bmod p-1$ and γ multiplication $\bmod n$.

Cryptanalysis

This system is insecure because the modulus n can be factored by the Pollard $p-1$ method [P]. This method is effective because $p-1$ is highly composite. To explain the basic idea of this factoring algorithm, let $d \geq \max\{d_i \mid 1 \leq i \leq k\}$.

Lemma 1. $p \mid gcd(a^{B!^d}-1, n)$.

Proof. $a^{p-1} \equiv 1 \bmod p$, hence $a^{\prod_{i=1}^{k} p_i^{d_i}} \equiv 1 \bmod p$. Since $\prod_{i=1}^{k} p_i^{d_i} \mid B!^d$; $a^{B!^d} \equiv a^{\prod_{i=1}^{k} p_i^{d_i}} \equiv 1 \bmod p$.□

The cryptanalyst will not know the values of B and d. He can just choose some B' and d'. If $gcd(a^{B'!^{d'}}-1, n)=1$, then he knows that he picked B' or d' too small, and he could just increase them, perhaps by doubling them. The gcd could also be n, but this would only happen if $q-1$ was also highly composite. In this case, the cryptanalyst could just modify the values of B' and d' until he found one that would give a gcd of either p or q. The expected running time for this cryptanalysis is $O(dB \log B \log n)$.

System 2

Let $a_0, a_1, \ldots, a_{k-1}$ be randomly chosen positive integers. Let A be the matrix

$$\begin{bmatrix} 1 & a_0 & a_0^2 & \dots & a_0^{k-1} \\ 1 & a_1 & a_1^2 & \dots & a_1^{k-1} \\ & & \vdots & & \\ 1 & a_{k-1} & a_{k-1}^2 & \cdots & a_{k-1}^{k-1} \end{bmatrix}$$

Given an integer x, with binary representation $x_0, x_1, \dots, x_{k-1}$, let $\bar{x} = (x_0, \dots, x_{k-1})$. The encryption of x is the vector $\bar{y} = A\bar{x}$. Decryption is preformed by multiplying by A^{-1}. This is a privacy homomorphism with γ being addition or subtraction over the integers, and α being componentwise addition or subtraction over the integers.

Cryptanalysis

Suppose j is the largest index such that $x_j = 1$, i.e., $x_{j+1}, \dots, x_{k-1} = 0$. Let z be any positive integer. Then $z^j \le \sum_{i=0}^{j} x_i z^i \le \sum_{i=0}^{j} z^i < (z+1)^j$. Thus $\left\lfloor \left(\sum_{i=0}^{j} x_i z^i \right)^{\frac{1}{j}} \right\rfloor = z$. This leads to an obvious ciphertext only attack. Given a cipher $y = (y_0, \dots, y_{k-1})$, guess a value for j, the largest index such that $x_j = 1$. Compute $\lfloor y_0^{1/j} \rfloor = b_0$. Write y_0 in base b_0 notation. If all the coefficients are 0 and 1, then probably $b_0 = a_0$ and x is easily found. If not try a different choice for j. The values of $y_1, \dots, y_{k-1}$ can be used as an additional check.

System 3

Let p and q be large primes. Encryption of message M is an ordered pair $(M \bmod p, M \bmod q)$. This is a privacy homomorphism with γ being addition, subtraction or multiplication componentwise, and α being addition, subtraction or multiplication $\bmod pq$. Decryption is done using the Chinese remainder theorem.

Cryptanalysis

We will argue that this can be broken by a known plaintext attack. Let the encryption of a message M be represented by a pair (C, D), where $C \equiv M \bmod p$ and $D \equiv M \bmod q$. Assume that the cryptanalyst has plaintext-ciphertext pairs M_i, C_i, D_i, $i = 1, 2, 3, \dots, r$. $p \mid C_i - M_i$. Let p' be the gcd $\{C_i - M_i : 1 \le i \le r\}$. Clearly, $p \mid p'$. A q' such that $q \mid q'$ can be found in a similar manner. If $p' = p$ and $q' = q$, the cryptanalyst can decrypt any ciphertext. Even for small r, there is a high probability that $p' = p$ and $q' = q$. Even if this is not the case, then for any new ciphertext the cryptanalyst can either determine the plaintext, or if he is given the plaintext, he can improve his knowledge of either p or q. Specifically, given ciphertext (C, D), the cryptanalyst can find M' such that $M' \equiv C$

$\bmod p'$ and $M' \equiv D \bmod q'$. If $M' \neq M$, then either $M \not\equiv C \bmod p'$ or $M \not\equiv D \bmod q'$. So if the cryptanalyst is given this M, he can improve either p' or q' by replacing p' by $gcd(M - C, p')$ and q' by $gcd(M - D, q')$. The cryptanalyst would have to be given at most $\log p + \log q$ such messages before he knows p and q.

System 4

In this system encryption is just writing the message in a secret radix system. Referring to the least significant digit, this reduces to the cryptanalysis of the previous system.

3. An R-additive privacy homomorphism

In this section, we introduce the notion of an R-additive privacy homomorphism in which only the addition of at most R messages is allowed. We will also give an example that appears to be secure against a ciphertext only attack. We will call this example the modular multiplication R-additive privacy homomorphism or MM-RAPH.

Let $\mathbf{M}$ be a subset of the integers. An R-additive privacy homomorphism is a family of functions $(e_k, d_k, \alpha, \gamma)$ such that $d_k(\gamma(e_k(m_1), \ldots, e_k(m_r)) = m_1 + \cdots + m_r$ for $r \leq R$ and for each k in key space $\mathbf{K}$ and any $m_1, \ldots, m_r$ in message space $\mathbf{M}$.

The [ALN] chosen ciphertext attack on privacy homomorphisms in which γ is addition will also apply to R-additive privacy homomorphisms if R is large enough. In particular, if C is the set of integers between 0 and 2^n, then the chosen ciphertext attack will work as follows. For $i = 0, 1, \ldots, n-1$, find m_i such that $e_k(m_i) = 2^i$. Given a cipher $c = \sum_{i=0}^{n-1} x_i 2^i$, for $x_i \in \{0, 1\}$, then $d_k(c) = \sum_{i=0}^{n-1} x_i m_i$ if $R > \sum_{i=0}^{n-1} x_i$. So if $R > n$, this attack will always work. However, if R is a small constant, then this attack would need an exponential number of chosen ciphertext-plaintext pairs to be successful.

Suppose that the messages that we actually want to encrypt are the integers between 0 and B. Let ℓ be a fixed positive integer. Let $N = RB$. Let z be an integer randomly chosen in the interval $[N^\ell, 2N^\ell]$, and let y be chosen relatively prime to z.

A block of ℓ messages $m_0, \ldots, m_{\ell-1}$ will be encrypted by computing $t = \sum_{i=0}^{\ell-1} m_i N^i$, and then computing $c = ty \bmod z$.

The decryption process is obvious. Given a cipher c form $t = cy^{-1} \bmod z$. Write t base N to obtain $m_0, \ldots, m_{\ell-1}$. Showing that this is an R-additive privacy homomorphism is straightforward as well.

We can modify System 3 of [RAD] to make an RAPH similar to the MM-RAPH. We merely choose p and q so that pq is in the interval $[N^\ell, 2N^\ell]$. Then to encrypt a block of messages $m_0, \ldots, m_{\ell-1}$ we again compute $t = \sum_{i=0}^{\ell-1} m_i N^i$, and then form the ordered pair $(t \bmod p, t \bmod q)$. The security of this system is similar to the security of the MM-RAPH, but we will not discuss the details here.

4. Attacks on the R-additive privacy homomorphism

The modular multiplication RAPH is immediately suspect to the attacks that broke the knapsack cryptosystems. However, it appears that these attacks will not be successful against the modular multiplication RAPH when z is large enough relative to R. We will not specify how large z must be, because it depends upon the performance of the Lovasz basis reduction algorithm [LLL], and there have not been enough computational experiments to estimate what this performance is.

For $a_1, \ldots, a_n, A, b_1, \ldots, b_n, B$ all positive integers, we will say that a vector $\boldsymbol{\beta} = \left(\frac{b_1}{B}, \ldots, \frac{b_n}{B} \right)$ is a δ-quality simultaneous Diophantine approximation (or just a δ-approximation) to $\alpha = \left(\frac{a_1}{A}, \ldots, \frac{a_n}{A} \right)$ if $B < A$ and

$\max_{1 \le i \le n} \mid B \frac{a_i}{A} - b_i \mid \le A^{-\delta}$. See [L].

For most tuples $a_1, \ldots, a_n, A$ there are no δ-approximations for fixed $\delta > \frac{1}{n}$. On the other hand, by Dirichlet's theorem, there is always a δ-approximation for fixed $\delta < \frac{1}{n}$, specifically if $A^{-\delta} > \left(\lceil A^{\frac{1}{n}} \rceil - 1 \right)^{-1}$.

Suppose now that we are given $n+1$ ciphers $c_0, c_1, \ldots, c_n$. The attack that we will examine involves finding simultaneous diophantine approximations to the vector $\left(\frac{c_1}{c_0}, \ldots, \frac{c_n}{c_0} \right)$. We will assume that c_0 is the largest of the c_i. If it is not, we would need to reduce the c_i mod c_0 and this would complicate our discussion. If the c_i were random, we would expect the best δ-approximation to be for $\delta \approx \frac{1}{n}$. However, these c_i are not random. In particular there exist integers k_i, such that $c_i\ y^{-1} - k_i\ z = t_i < \frac{z}{R}$. This implies that

$$\frac{y^{-1}}{z} - \frac{k_i}{c_i} = \frac{t_i}{c_i z}$$

$$\frac{k_0}{c_0} - \frac{k_i}{c_i} = \frac{t_i}{c_i z} - \frac{t_0}{c_0 z}$$

$$k_0 \frac{c_i}{c_0} - k_i = \frac{t_i}{z} - \frac{t_0\, c_i}{z\, c_0}$$

$$\mid k_0 \frac{c_i}{c_0} - k_i \mid \ < \frac{1}{R}\ . \tag{4.1}$$

If $\ell > 1$, there are other approximations that are of about this same quality. The reason for these other approximations is similar to the reason for other approximations existing in the multiple iterated knapsacks, [B], but we will not go into the details here. We will simply indicate that these other approximations exist because each $t_j = \sum_{i=0}^{\ell-1} m_{ji}\ N^i$, and $m_{ji} < N/R$. If we had only $m_{ji} < N$, these other approximations would not exist. It can be shown that if ℓsuch interesting approximations can be found then the privacy homomorphism can be broken.

In n is large enough so that $\frac{1}{R} << c_1^{\frac{-1}{n}}$, i.e. $R^n >> c_1$, then the interesting approximations will probably be best approximations, but only by a factor of $\frac{1}{R}$. So the security of this system depends on whether we can find these approximations.

The Lovasz algorithm can be used to find good simultaneous diophantine approximations. It is guaranteed to find a δ-approximation within a factor of 2^n of the best approximation for a given vector of integers. Computational experience has shown that it often does much better. The attacks on some knapsack systems relied on the Lovasz algorithm finding approximations that are within n of the best approximations. However computational experience indicates that the Lovasz algorithm is unlikely to find approximations that are only within a constant of the best. Further computational experiments on our privacy homomorphism will be necessary to conclude that the Lovasz algorithm will not break it.

5. A modification to the MM-RAPH

In this section, we discuss a modification to the MM-RAPH that appears to increase security, although it is still susceptible to a known plaintext attack.

Again, let the integers between 0 and B be the messages that we want to encrypt. Let $N = R\,B$. Let p be a prime larger than N. (The first prime larger than N will be satisfactory.) Let A be an ℓ by ℓ matrix chosen randomly such that A is nonsingular $\bmod p$. Let q be an integer $>Rp$. Let z be an integer randomly chosen in the interval $[q^{\ell}, 2q^{\ell}]$ and let y be chosen relatively prime to z.

A block of ℓ messages $m_0, \ldots, m_{\ell-1}$ will be encrypted by taking the vector $\bar{m} = (m_0, \ldots, m_{\ell-1})$ and forming $\bar{s} = A\,\bar{m} \bmod p$. Let $\bar{s} = (s_0, \ldots, s_{\ell-1})$. Compute $t = \sum_{i=0}^{\ell-1} s_i\, q^i$. Finally set $c = ty \bmod z$.

The decryption process is obvious. Given a cipher c form $t = cy^{-1} \bmod z$. Write t base q to obtain $\bar{s} = (s_0, \ldots, s_{\ell-1})$. Then $A^{-1}\,\bar{s} \bmod p$ gives the desired message $\bar{m}$.

Showing that this is an R-additive privacy homomorphism is straightforward as well. Suppose $\bar{m}_1, \ldots, \bar{m}_r$ are r messages for $r \le R$. Let $\bar{s}_j = (s_{j,0}, \ldots, s_{j,\ell-1})$, t_j, and c_j be the values computed in the encryption of $\bar{m}_j$. We need to show that when $c = c_1 + \cdots + c_r$ is decrypted the result is $\bar{m} = (m_{1,0} + \cdots + m_{r,0}, \ldots, m_{1,\ell-1} + \cdots + m_{r,\ell-1})$.

Clearly $t = cy^{-1} \equiv t_1 + \cdots + t_r \bmod z$. But $\sum_{j=1}^{r} t_j = \sum_{j=1}^{r} \sum_{i=0}^{\ell-1} s_{j,i}\, q^i = \sum_{i=0}^{\ell-1} \left[\sum_{j=1}^{r} s_{j,i} \right] q^i$. Also $\sum_{j=1}^{r} s_{j,i} < q$, for $0 \le i \le \ell-1$. So $\sum_{j=1}^{r} t_j < z$ and writing t base q gives the values $\sum_{j=1}^{r} s_{j,i}$, for $0 \le i \le \ell-1$. Let $\bar{s} = \left[\sum_{j=1}^{r} s_{j,0}, \ldots, \sum_{j=1}^{r} s_{j,\ell-1} \right]$. Then $\bar{s}\,A^{-1} \equiv \bar{m}_1 + \cdots + \bar{m}_r \bmod p$. But since $m_{1,i} + \cdots + m_{r,i} < p$, for $0 \le i \le \ell-1$, the result follows.

We will not give all of the details of why this modified system is still susceptible to a known plaintext attack. The basic idea is that known plaintext-ciphertext pairs can be used to produce vectors that have simultaneous Diophantine approximations that are better than the approximations described in (4.1).

6. Open problems

We have introduced an R-additive privacy homomorphism for which we have argued that it cannot be broken by an obvious application of the attacks that broke the knapsack cryptosystems. Is this scheme in fact secure? It would be interesting to find other privacy homomorphisms.

References

[ALN] N. Ahituv, Y. Lapid, S. Neumann, "Processing Encrypted Data", CACM, Sept. 1987, Vo. 30, No. 9, pp. 777-780.

[B] E. F. Brickell, ''Breaking Iterated Knapsacks'', *Advances in Cryptology, Proc. Crypto 84,* Santa Barbara, August 19-22, 1984, Lecture Notes in Computer Science, vol. 196, Springer-Verlag, Berlin, 1985, pp. 342-358.

[BLY] E. F. Brickell, P. J. Lee, Y. Yacobi, "Secure Audio Teleconference", to appear in *Advances in Cryptology, Proc. Crypto 87,* Springer-Verlag, New York.

[L] J. C. Lagarias, "Knapsack Public Key Cryptosystems and Diophantine Approximation" (Extended Abstract), *Advances in Cryptology, Proc. Crypto 83*, Plenum Publ. Co., New York, 1984, pp.3-24.

[LLL] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovasz, ''Factoring Polynomials with Rational Coefficients'', *Mathematische Annalen 261*, pp. 515-534, 1982.

[PH] Stephen C. Pohlig and Martin E. Hellman, ''An Improved algorithm for computing Logarithms over $GF(p)$ and its cryptographic significance'', IEEE Trans. on Inf. Th. Vol. IT-24, No. 1 Jan. 1978. pp. 106-110.

[P] J. M. Pollard, ''Theorems on factorization and primality testing'', Proc. Cambridge Philos. Soc. vol. 76 (1974), pp. 521-528.

[RAD] Ronald L. Rivest, Len Adleman and Michael L. Dertouzos, ''On data banks and privacy homomorphisms'', in Foundation of Secure Computations, Academic Press 1978.

[RSA] R. L. Rivest, A. Shamir and L. Adleman, ''A Method for Obtaining Digital Signatures and Public Key Cryptosystems'', *Commun. ACM,* vol. 21, pp. 294-299, April 1978.

[S] A. Shamir, ''A Polynomial Time Algorithm for Breaking the Basic Merkle-Hellman Cryptosystem'', IEEE Trans. Inform. Theory, vol. IT-30, No. 5, September 1984, pp. 699-704.