A new security notion for asymmetric encryption Draft #12

Muhammad Rezal Kamel Ariffin[1,2]

[1] Al-Kindi Cryptography Research Laboratory,
Institute for Mathematical Research,
[2] Department of Mathematics, Faculty of Science,
Universiti Putra Malaysia (UPM), Selangor, Malaysia
rezal@upm.edu.my

Abstract. A new practical asymmetric design is produced with desirable characteristics especially for environments with low memory, computing power and power source.

KEYWORDS: asymmetric security, diophantine equations, bivariate function hard problem

1 Introduction

In this article we provide a new asymmetric encryption design based on the difficulty of solving ***solving a diophantine equation with infinitely many solutions*** and ***solving a system of diophantine equations with unknown exponent***. Further discussion on this problem will be provided in the following sections.

2 A new security notion for asymmetric encryption

The following 2 sub-sections provide definitions and discussion on the the so-called ***underlying security primitive*** which the our asymmetric scheme relies on.

2.1 Linear diophantine equations with infinitely many solutions

Definition 1. *To determine the preferred solution for a diophantine equation where that preferred solution is from a set of infinitely many solutions.*

To further understand and obtain the intuition of Definition 1, we will now observe a remark by Herrmann and May [1]. It discusses the ability to retrieve variables from a given linear Diophantine equation. But before that we will put forward a famous theorem of Minkowski that relates the length of the shortest vector in a lattice to the determinant[1]:

Theorem 1. *In an ω-dimensional lattice, there exists a non-zero vector v with*

$$||v|| \leq \sqrt{\omega}\det(L)^{\frac{1}{\omega}}$$

We now put forward the remark.

Remark 1. There is a method for finding small roots of linear modular equations $a_1x_1 + a_2x_2 + ... + a_nx_n \equiv 0 \pmod{N}$ with known modulus N. It is further assumed that $\gcd(a_i, N) = 1$. Let X_i be upper bound on $|x_i|$. The approach to solve the linear modular equation requires to solve a shortest vector problem in a certain lattice. We assume that there is only one linear independent vector that fulfills the Minkowski bound (Theorem 1) for the shortest vector. Herrmann and May showed that under this heuristic assumption that the shortest vector yields the unique solution $(y_1, ..., y_n)$ whenever

$$\prod_{i=1}^{n} X_i \leq N.$$

If in turn we have

$$\prod_{i=1}^{n} X_i > N^{1+\epsilon}.$$

then the linear equation usually has N^ϵ many solutions, which is exponential in the bit-size of N. So there is no hope to find efficient algorithms that in general improve on this bound, since one cannot even output all roots in polynomial time.

We now put forward a corollary.

Corollary 1. *A linear diophantine equation $f(x_1, x_2, ..., x_n) = a_1x_1 + a_2x_2 + ... + a_nx_n = N$, with*

$$\prod_{i=1}^{n} x_i > N^{1+\epsilon}.$$

is able to ensure secrecy of the preferred sequence $\mathbf{x} = \{x_i\}$.

Remark 2. In fact if one were to try to solve the linear diophantine equation $N = a_1x_1 + a_2x_2 + ... + a_nx_n$, where $\prod_{i=1}^{n} X_i > N^{1+\epsilon}$, any method will first output a short vector $\mathbf{x} = \{x_i\}$ as the initial solution. Then there will be infinitely many values from this initial condition that is able to recontruct N.

2.2 System of diophantine equations with unknown exponent(s) and reduction moduli

It is well known that from:

$$A \equiv g^a \pmod{p}$$

if given the tuple (A, g, p) to determine the unknown exponent a (if the tuple are "strong") would be difficult. In fact this is the discrete log problem (DLP).

We now extend this feature to the following setting; given:

$$A_i \equiv \sum_{j=1}^{k} g_j^{a_j} \pmod p$$

If given the tuple (A_i), determine (a_j, g_j, p).

3 Bivariate Function Hard Problem (BFHP)

In this section we introduce a particular case of a linear diophantine equation in 2 variables that is able to secure its private parameters under some conditions. This section explores subsection 2.1 in more detail for the mentioned case.

Definition 2. *We define $\mathbb{Z}^+_{(2^{m-1},2^m-1)}$ as a set of positive integers in the interval $(2^{m-1}, 2^m - 1)$. In other words, if $x \in (2^{m-1}, 2^m - 1)$, x is an m-bit positive integer.*

Proposition 1. *Let $A = f(x_1, x_2, ..., x_n)$ be a one-way function that maps $f : Z^n \to Z^+_{(2^{m-1},2^m-1)}$. Let f_1 and f_2 be such function (either identical or non-identical) such that $A_1 = f(x_1, x_2, ..., x_n), A_2 = f(y_1, y_2, ..., y_n)$ and $gcd(A_1, A_2) = 1$. Let $u, v \in Z^+_{(2^{n-1},2^n-1)}$. Let (A_1, A_2) be public parameters and (u, v) be private parameters. Let*

$$G(u, v) = A_1u + A_2v \tag{1}$$

with the domain of the function G is $Z^2_{(2^{n-1},2^n-1)}$ since the pair of positive integers $(u, v) \in Z^2_{(2^{n-1},2^n-1)}$ and $Z^+_{(2^{m+n-1},2^{m+n}-1)}$ is the codomain of G since $A_1u + A_2v \in Z^+_{(2^{m+n-1},2^{m+n}-1)}$.

If at minimum $n - m - 1 = k$, where 2^k is exponentially large for any probabilistic polynomial time (PPT) adversary to sieve through all possible answers, it is infeasible to determine (u, v) over $\mathbb{Z}$ from $G(u, v)$. Furthermore, (u, v) is unique for $G(u, v)$ with high probability.

Before we proceed with the proof of the above proposition we would like to put forward 2 remarks.

Remark 3. We remark that the preferred pair $(u, v) \in \mathbb{Z}$, is the prf-solution for (1). The preferred pair (u, v) is one of the possible solutions for (1) from:

$$u = u_0 + A_2t \tag{2}$$

and

$$v = v_0 - A_1t \tag{3}$$

for any $t \in \mathbb{Z}$.

Remark 4. Before we proceed with the proof, we remark here that the diophantine equation given by $G(u, v)$ is solved when the preferred parameters $(u, v) \in \mathbb{Z}$ are found. That is the BFHP is prf-solved when the preferred parameters $(u, v) \in \mathbb{Z}$ are found.

Proof. We begin by proving that (u, v) is unique for each $G(u, v)$ with high probability. Let $u_1 \neq u_2$ and $v_1 \neq v_2$ such that

$$A_1u_1 + A_2v_1 \neq A_1u_2 + A_2v_2 \tag{4}$$

We will then have

$$Y = v_2 - v_1 = \frac{A_1(u_1 - u_2)}{A_2}$$

Since $\gcd(A_1, A_2) = 1$ and $A_2 \approx 2^n$, then the probability that Y is an integer is 2^{-n}. Then the probability that $v_1 - v_2$ is an integer solution not equal to zero is 2^{-n}. Thus $v_1 = v_2$ with probability $1 - 2^{-n}$.

We next proceed to prove that to prf-solve the diophantine equation given by (1) is infeasible. The general solution for $G(u, v)$ is given by (2) and (3) for some integer t.

To find u within the stipulated interval $u \in (2^{n-1}, 2^n - 1)$ we have to find the integer t such that the inequality $2^{n-1} < u < 2^n - 1$ holds. This gives

$$\frac{2^{n-1} - u_0}{A_2} < t < \frac{2^n - 1 - u_0}{A_2}$$

Then, the difference between the upper and the lower bound is $\approx \frac{2^{n-2}}{2^m}$.

Since $n - m - 1 = k$ where 2^k is exponentially large for any probabilistic polynomial time (PPT) adversary to sieve through all possible answers, we conclude that the difference is very large and finding the correct t is infeasible. This is also the same scenario for v.

Example 1. Let $A_1 = 191$ and $A_2 = 229$. Let $u = 41234$ and $v = 52167$. Then $G = 19821937$. Here we take $m = 16$ and $n = 8$. We now construct the parametric solution for this BFHP. The initial points are $u_0 = 118931622$ and $v_0 = -99109685$. The parametric general solution are: $u = u_0 + A_2t$ and $v = v_0 - A_1t$. There are approximately $286 \approx 2^9$ (i.e. $\frac{2^{16}}{229}$) values of t to try (i.e. difference between upper and lower bound), while at minimum the value is $t \approx 2^{16}$. In fact, the correct value is $t = 519172 \approx 2^{19}$.

4 A new asymmetric primitive

In this section we provide the reader with a working cryptographic primitive discussed in subsection 2.2 and section 3.

• Key Generation by Along

INPUT: The size n of the parameters.
OUTPUT: A public key tuple (n, e_1, e_2, e_3) and private keys (d_1, d_2, p).

1. Generate 3 private random n-bit prime,(p, p_1, p_2).
2. Generate e where $gcd(e, p-1) = 1$. For reasons to be observed later the value of e is with reference to the amount of data the user intends to relay.
3. Compute private $d_1 \equiv e^{-1} \pmod{p-1}$.
4. Generate secret random (preferably primes) n-bit $g_1, k_1, k_2, t_1, t_2 \in \mathbb{Z}_p$.
5. Compute secret $A = p_1 - 2k_1 + t_1$, $B = p_2 - 2k_2 + 2t_2$, $\alpha = g_1^A \pmod{p}$ and $d_B \equiv B^{-1} \pmod{p-1}$. Observe that B is always odd. Hence, d_B will exist.
6. Compute secret $g_2 \equiv \alpha^{d_B} \pmod{p}$. We now have $g_1^A \equiv g_2^B \pmod{p}$.
7. Compute secret $T \equiv g_1^{p_1 - t_1} - g_2^{p_2 - t_2} \pmod{p}$.
8. Compute secret $a_1 \equiv T^{-1} \pmod{p}$ and $a_2 \equiv 2T^{-1} \pmod{p}$.
9. Compute public values
   (a) $e_1 \equiv g_1^{k_1 - t_1} \pmod{p}$,
   (b) $e_2 \equiv g_2^{k_2 - t_2} \pmod{p}$,
   (c) $e_3 \equiv a_1 g_1^{p_1 - k_1} + a_2 g_2^{p_2 - k_2} \pmod{p}$.
10. Compute private key $d_2 \equiv a_1 g_2^{p_2 - k_2} + a_2 g_1^{p_1 - k_1} \pmod{p}$.
11. Return the public key tuple (n, e_1, e_2, e_3) and private key tuple (d_1, d_2, p).

• Encryption by Busu

INPUT: Along's public key set (n, e_1, e_2, e_3) and the message M tuple (b_0, b_1, b_3) where $b_3 \approx 2^{n-1}$ and $b_0, b_1 \approx 2^{(e-2)n}$.
OUTPUT: A ciphertext pair (C_1, C_2).

1. Compute $b_2 = b_3^e - b_0 + 2b_1$.
2. Compute the first ciphertext $C_1 = b_0 + b_1(e_2 e_3) + b_2(e_1 e_3)$.
3. Compute the second ciphertext $C_2 = b_1 e_1 + b_2 e_2$.
4. Send the ciphertext pair $C = (C_1, C_2)$.

• Decryption by Along

INPUT: The ciphertext pair $C = (C_1, C_2)$ and private key tuple (d_1, d_2, p).
OUTPUT: The message tuple $M = (b_0, b_1, b_3)$.

1. Compute $b_3 \equiv (C_1 - C_2 d_2))^{d_1} \pmod{p}$.
2. Solve the simultaneous equations $C_1 - b_3^e = b_1(e_2 e_3 + 2) + b_2(e_1 e_3 - 1)$ and $C_2 = b_1 e_1 + b_2 e_2$ to obtain (b_1, b_2). Then obtain $b_0 = 2b_1 - b_2 + b_3^e$.
3. Return the message tuple $M = (b_0, b_1, b_3)$.

Proposition 2. *The decryption process is correct.*

Proof. From $g_1^A - g_2^B \equiv 0 \pmod p$ and $T \equiv g_1^{p_1-t_1} - g_2^{p_2-t_2} \pmod p$, we have

$$\begin{aligned}(C_1 - C_2 d_2))^{d_1} &\equiv (b_0 + (a_1 b_2 - a_2 b_1)T + (g_1^{k_1-t_1} - g_2^{k_2-t_2})(g_1^A - g_2^B))^{d_1} \\ &\equiv (b_0 - 2b_1 + b_2)^{d_1} \\ &\equiv (b_3^e)^{d_1} \\ &\equiv b_3 \pmod p.\end{aligned}$$

We obtain the exact b_3 since $b_3 < p$, which ensures that no modular reduction has occurred. Next, to obtain (b_0, b_1) is trivial.

In the next section we will point out locations where the fundamental source of security situated.

5 The fundamental source of security

We will dissect the mathematical structures introduced in the above so-called "cryptosystem". We will begin at looking at Along's parameters first.

5.1 Security of the ciphertext

- Observe the ciphertext given by $C_1 = b_0 + b_1(e_2e_3) + b_2(e_1e_3)$. We have $C_1 \approx 2^{(e+2)n}$ while $b_0b_1b_2 \approx 2^{(2e+1)n}$. Thus, $b_0b_1b_2 > C_1$.
- We have $b_1, b_2 \approx 2^{en}$ while $e_2, e_3 \approx 2^n$, thus the equation $C_2 = b_1e_1 + b_2e_2$ is "protected" by BFHP.
- We also have $C_2 \approx 2^{(e+1)n}$ while $b_1b_2 \approx 2^{2en}$. Thus, $b_1b_2 > C_2$.
- To solve the simultaneous equations of C_1, C_2 it is a system of 2 equations with 3 variables.

5.2 Security of the public key

In this section we study methods to evolve the equations behind the public-private key equations such that one can obtain the equation $y \equiv 0 \pmod p$. That is, to obtain an equation that would lead to a factorization problem (i.e which is not the objective of our research).

Security type-1
The public-private key of the scheme in this article follows the following system of equations:

$$a_1T - 1 \equiv 0 \pmod p \tag{5}$$
$$a_2T - 2 \equiv 0 \pmod p \tag{6}$$
$$\delta_1 e_1 - \delta_2 e_2 \equiv 0 \pmod p \tag{7}$$

where:

$$T \equiv g_1^{p_1 - t_1} - g_2^{p_2 - t_2} \pmod{p} \tag{8}$$
$$\delta_1 \equiv g_1^{p_1 - 3k_1 + 2t_1} \pmod{p} \tag{9}$$
$$\delta_2 \equiv g_2^{p_2 - 3k_2 + 2t_2} \pmod{p} \tag{10}$$

This is a system of 3 equations with 11 variables $(p, p_1, p_2, a_1, a_2, g_1, g_2, t_1, t_2, k_1, k_2)$.

6 Subset sum - like problem?

To obtain parameters that satisfy equations (5)-(10) simultaneously "mimics" the subset sum problem.

7 Data overhead

Both ciphertexts are in total $\approx (2e - 1)n$-bits. While for the message tuple $M = (b_0, b_1, b_3)$ is $(2e - 3)n$-bits. The message expansion is $\approx 2n$-bits.

8 Collision type attacks

We dedicate this section to discuss the possibility of designing a collision type attack on our new scheme.

9 Achieving IND-CCA2

It is obvious that the new scheme achieves IND-CPA. But how about IND-CCA2?

10 Conclusion

This paper presents a new cryptosystem that has advantages in the following areas against known public key cryptosystems:

1. It has a complexity order of $O(n^2)$ during encryption and $O(n^3)$ during decryption.
2. *Mathematically, an adversary does not have any advantage to attack the published public key or the ciphertext.*
3. Does the new scheme produce "cylic-type" features that would allow a collision type attack to be designed?
4. If a collision type attack cannot be designed, how do we propose to evaluate the scheme in order to suggest a minimum key length?

11 Acknowledgment

I would like to acknowledge Prof Abderrahmane Nitaj of University of Caen, France and Dr Yanbin Pan of the Chinese Academy of Science, China for all discussion and feedbacks while designing the scheme.

References

1. Herrmann, M., May, A.: Solving Linear Equations Modulo Divisors: On Factoring Given Any Bits. In ASIACRYPT 2008, volume 5350 of Lecture Notes in Computer Science, pp. 406-424. Springer-Verlag (2008)