Cryptanalysis and Improvement of Akleylek et al.'s cryptosystem

Roohallah Rastaghi

Department of Electrical Engineering, Aeronautical University of Science and Technology, Tehran, Iran.

r.rastaghi59@gmail.com

Keywords: Cryptography, Cryptanalysis, ElGamal Cryptosystem, Knapsack problem.

Abstract. Akleylek et al. [S. Akleylek, L. Emmungil and U. Nuriyev, Algorithm for peer-to-peer security, *journal of Appl. Comput. Math.*, Vol. 6(2), pp.258-264, 2007.], introduced a modified algorithm with steganographic approach for security in peer-to-peer (P2P) network. In this cryptosystem, Akleylek et al. attempt to increase the security of P2P network by connecting the ElGamal cryptosystem with knapsack problem. We show that this combination leak the security and makes the hybrid cryptosystem vulnerable to *ciphertext only attack*. Thus, in the network, an attacker can apply this attack and simply can recover the original message (plaintext) from any *challenge ciphertext*. Moreover, we show that the receiver cannot decrypt the ciphertext in polynomial time and so, the proposed cryptosystem is completely impractical. We modify this cryptosystem to increase security and efficiency.

1 Introduction

The use of computer network is increased day by day. This increment causes the number of nodes to increase. By increasing the client, the server becomes busy and insufficient although the bandwidths are high enough. Moreover, since the variety of requests is increased, servers may not have data the user needs. We can overcome these obstacles by using peer-to-peer (P2P) network. The P2P network did not have centralized server, some powerful nodes act as servers. In the fourth generation, streams over P2P network are supported. So each node can talk with another. The most important problem in the P2P network is management and security. There are several ways to make P2P networks secure. Cryptography has the most important role in each way. Cryptography is the art of keeping the data secure from eavesdropping and other malicious activities. Therefore, cryptographic algorithms are very useful in P2P systems since they can simultaneously protect message for an individual, and verify its integrity.

Akleylek et al. [1], introduced a modified algorithm with steganographic approach for security in P2P networks. In this cryptosystem, Akleylek et al. attempt to increase the security of P2P system by connecting the ElGamal cryptosystem [2] with the knapsack problem. The knapsack problem is a decision problem, which is NP-complete [3,5,6]. That is to say, this problem cannot be easily solved even using quantum computers. They use the ElGamal scheme to disguise private knapsack (easy knapsack) in order to produce public-key (hard knapsack). But as we show, this combination leaks the security and makes the cryptosystem vulnerable to ciphertext only attack. Hence, in the network an attacker can apply this attack and simply can recover plaintext from any "challenge ciphertext". In addition, we show that this cryptosystem is impractical. We try to modify it for increasing security and efficiency.

The rest of this paper is organized as follows: In the next section we give some mathematical background. Akleylek et al. cryptosystem will be presented in Section 3. Cryptanalysis of this cryptosystem will be discussed in Section 4 and in Section 5, we modify this cryptosystem to achieve the desired security and efficiency. Some conclusion is given in Section 6.

2 Preliminaries

In this section, we give some mathematical background and definitions which are needed to demonstrate our attack.

2.1 Mathematical background

Definition 1 (Subset sum problem[1]). *Given a set of positive integers* $(a_1, \ldots, a_n)$ *and a positive integer* s*. Whether there is a subset of the* a_i*'s such that their sum equal to* s*. That is equivalent to determine whether there are variables* $(x_1, \ldots, x_n)$ *such that*

$$s = \sum_{i=1}^{n} a_i x_i, \quad x_i \in \{0, 1\}, \quad 1 \leq i \leq n.$$

The subset sum problem is a decision problem, which is NP-complete. The computational version of the subset sum problem is NP-hard [5]

Definition 2 (super-increasing sequence). *The sequence* $(a_1, \ldots, a_n)$ *of positive integers is a super-increasing sequence, if* $a_i > \sum_{j=1}^{i-1} a_j$ *for all* $i \geq 2$.

[1] Additive knapsack problem

There is an efficient greedy algorithm to solve the subset sum problem if the a_i's are a super-increasing sequence: Just subtract the largest possible value from s and repeat. The following algorithm efficiently solves the subset sum problem for super-increasing sequences in the polynomial time.

Algorithm 1[5] Solving a super-increasing subset sum problem.

Input: Super-increasing sequence $(a_1, \ldots, a_n)$ and an integer s which is the sum of a subset of the a_i.
Output: $(x_1, \ldots, x_n)$ where $x_i \in \{0, 1\}$, such that $s = \sum_{i=1}^{n} a_i x_i$.

1. $i \leftarrow n$
2. While $i \geq 1$ do the following:
   (a) If $s \geq a_i$, then $x_i \leftarrow 1$ and $s \leftarrow s - a_i$. Otherwise $x_i \leftarrow 0$.
   (b) $i \leftarrow i - 1$.
3. Return $(x_1, x_2, \ldots, x_n)$.

Definition 3 (Subset product problem[2]). *A set of positive integers $(a_1, \ldots, a_n)$ and a positive integer d are given. Whether there exist a subset of the a_i's such that their product equals to d. That is equivalent determine whether there are variables $(x_1, \ldots, x_n)$ such that*

$$d = \prod_{i=1}^{n} a_i^{x_i}, \quad x_i \in \{0, 1\}, \quad 1 \leq i \leq n.$$

The Subset product problem is a decision problem, which is NP-complete [5]. As observed in [4,6], if the a_i's are small primes and much smaller than d, this problem can be solved in polynomial time by factoring d. Their result can be summarized in the following lemma.

Lemma 1. *If $(a_1, \ldots, a_n)$ are small primes, then we can solve the subset product problem in polynomial time.*

Proof. Since the a_i's are small primes and $x_i \in \{0, 1\}$, so we have

$$x_i = \begin{cases} 1 & \text{if } \mathsf{gcd}(d, a_i) = a_i \\ 0 & \text{if } \mathsf{gcd}(d, a_i) = 1 \end{cases}, \quad 1 \leq i \leq n$$

Hence

$$x_i = \begin{cases} 1 & \text{if } \quad a_i \mid d \\ 0 & \text{if } \quad a_i \nmid d \end{cases}, \quad 1 \leq i \leq n$$

where gcd means the greatest common divisor. Note that d is the product of distinct primes a_i, $1 \leq i \leq n$.

[2] Multiplicative knapsack problem

Definition 4 (*Discrete logarithm problem* (DLP)). *Given a prime p, a generator α of $\mathbb{Z}_p^*$, and an element $\beta \in \mathbb{Z}_p^*$. Find integer $x, 0 \leq x \leq p-2$, such that*

$$\alpha^x = \beta \mod p.$$

is called the discrete logarithm problem.

2.2 The ElGamal cryptosystem

The ElGamal cryptosystem is a public key cryptosystem based on the discrete logarithm problem in $(\mathbb{Z}_p^*, .)$. Let p be a large prime such that the DLP in $(Z_p^*, .)$ is infeasible, and let $g \in \mathbb{Z}_p^*$ be a primitive element. Each user selects a random integer a, $1 \leq a \leq p-2$, and compute $\beta = g^a \mod p$. $\{p, \alpha, \beta\}$ is public key and a is private key.

Suppose that we wish to send a message x to receiver. First, we select a random integer k such that $1 \leq k \leq p-2$. Then we compute $c_1 = \alpha^k \mod p$ and $c_2 = x.\beta^k \mod p$. We send ciphertext (c_1, c_2) to the receiver. The encryption operation in the ElGamal cryptosystem is randomize, since the ciphertext depends on both the plaintext x and on the random value k chosen by user. To recover plaintext x from ciphertext c, receiver uses the private-key a and compute $x = c_2(c_1^a)^{-1} \mod p$.

2.3 Ciphertext-only attack

A ciphertext-only attack is a scenario by which the adversary (or cryptanalyst) tries to deduce the decryption key by only observing the ciphertexts or decrypt a *challenge ciphertext*.

Attacker knowledge: some $y_1 = \mathsf{Enc}(x_1, pk)$, $y_2 = \mathsf{Enc}(x_2, pk)$, $\ldots$.

Attacker goal: obtain $x_1, x_2, \ldots$ or the secret-key sk.

Any encryption scheme vulnerable to this type of attacks is considered to be completely insecure.

3 Akleylek et al. Cryptosystem

In this section, we present Akleylek et al. cryptosystem. The authors intend to increase security of the proposed scheme by connecting the ElGamal cryptosystem with the knapsack problem.

3.1 Key generation

(a) Each user chooses a super-increasing sequence, $(a_1, \dots, a_n)$, such that $a_i > \sum_{i=1}^{j-1} a_i$, $2 \le j \le n$, and all a_i's are integer.

(b) The keys of ElGamal cryptosystem $\{\beta, g, p, a\}$ are calculated.

(c) For calculating public knapsack $pk = (b_1, \dots, b_n)$, randomly select an integer k, $1 \le k \le p-1$ and use the following operations:

$\beta = g^a \mod p$, $\quad s_i = g^k \mod p$, $\quad u_i = \beta^k.a_i \mod p$, $\quad$ and $b_i = (s_i, u_i)$, for $1 \le i \le n$.

Finally, the public-key $pk = ((s_1, u_1), \dots, (s_n, u_n))$ and the private-key $sk = \{\beta, g, p, a, (a_1, \dots, a_n)\}$ is obtained.

Remark 1. Note that Component $s_i = g^k \mod p$ of the public-key pk is constant respect to $i, 1 \le i \le n$.

3.2 Encryption

To encrypt n-bit binary message $\mathbf{x} = (x_1, \dots, x_n)$, we compute

$$c = (c_1, c_2) = \prod_{i=1}^{n} (s_i, u_i)^{x_i}. \tag{1}$$

We send ciphertext c to the receiver.

3.3 Decryption

To decrypt the ciphertext c, the receiver firstly calculates

$$d = c_2.(c_1^{-1})^a \mod p = \frac{\prod_{i=1}^{n} (u_i)^{x_i}}{\prod_{i=1}^{n} (s_i^a)^{x_i}} \mod p = \prod_{i=1}^{n} a_i^{x_i} \mod p. \tag{2}$$

Note that $u_i = \beta^k.a_i \mod p = g^{ka}.a_i \mod p = (s_i^a).a_i \mod p$.

After calculating d, we must obtain plaintext $\mathbf{x} = (x_1, \dots, x_n)$ from $d = a_1^{x_1}.a_2^{x_2}.\dots.a_n^{x_n}$.

3.4 A note about Akleylek et al. cryptosystem

From equation 2, we have $d = \prod_{i=1}^{n} a_i^{x_i}$ where $a_1, \dots, a_n$ is a super-increasing sequence. From Lemma 1, when a_i's are small primes, we can calculate x_i's from d, otherwise, the problem remains NP-complete and we cannot solve this problem. Here, since a_i's are super-increasing sequence, we cannot obtain $x_1, \dots, x_n$ from equation 2, in practice and so, Akleylek et al. cryptosystem is completely impractical.

4 Cryptanalysis of Akleylek et al. cryptosystem

In this section, we show that Akleylek et al.'s cryptosystem is vulnerable to ciphertext-only attack. In other words, we can obtain plaintext from any challenge ciphertext.

Suppose $c = (c_1, c_2)$ be any challenge ciphertext which encrypted with Akleylek et al.'s cryptosystem and we intend to find the corresponding plaintext. From equation 1, we have $c = (c_1, c_2) = \prod_{i=1}^{n} (s_i, u_i)^{x_i} = (s_1, u_1)^{x_1} \dots (s_n, u_n)^{x_n}$. The component $s_i = g^k \mod p$ of the public-key is constant for each i and we can assume $s_i = t$, $1 \le i \le n$. We have

$$c_1 = \prod_{i=1}^{n} s_i^{x_i} = t^{\sum_{i=1}^{n} x_i} = t^h, \tag{3}$$

where $h = \sum_{i=1}^{n} x_i$ is the Hamming weight (the number of $x_i = 1$) of the binary message $\mathbf{x} = (x_1, \dots, x_n)$. From equation 3, we can compute Hamming weight h of plaintext $x_1, \dots, x_n$ and so, we know the number of the x_i's where $x_i = 1$. From equation 1, we have $c_2 = \prod_{i=1}^{n} u_i^{x_i}$ and so, we know the number of u_i's where product of them equals to c_2, but we do not know which of them. For obtaining these u_i's, we need to find a h-tuple subset of $u_1, \dots, u_n$ from public-key $((*, u_1), \dots, (*, u_n))$ such that product of them equal to c_2. We denote this subset by S. We can choose h elements of $u_1, \dots, u_n$ in $\binom{n}{h}$ ways. So, we need at most $\binom{n}{h}$ bit operations to find such subsets. After obtaining these u_i's, we can obtain original plaintext from the following equation

$$x_i = \begin{cases} 1 & \text{if} \quad u_i \in S \\ 0 & \text{if} \quad u_i \notin S \end{cases} \qquad 1 \le i \le n.$$

We have

$$\binom{n}{h} = \frac{n(n-1)\dots(n-h+1)}{h(h-1)\dots 1} < \frac{n^h}{h!} < n^h.$$

Hence, the complexity of attack is $O(n^h)$ and polynomial time.

5 Modified cryptosystem

This cryptosystem is based on multiplicative knapsack problem. The ciphertext is obtained by multiplying the public-keys indexed by the message bits and the plaintext is recovered by factoring the ciphertext raised to a secret power.

(1) Key generation [7] Each user

(a) Choose large prime p such that discrete logarithm problem in $(\mathbb{Z}_p^*, .)$ is infeasible.

(b) Determine the largest integer n such that $p > \prod_{i=1}^{n} p_i$, where p_i is the i-th prime (start from $p_1 = 2$).

(c) Randomly choose integer a, k such that $1 < a, k < p-1$ and compute

$$\begin{aligned} \beta &= g^a && \mod p, \\ s_i &= g^k && \mod p, \\ u_i &= \beta^k.p_i && \mod p, \end{aligned}$$

and $b_i = (s_i, u_i)$ for $1 \leq i \leq n$. $\{n, p, (b_1, \ldots, b_n)\}$ is the public-key and $\{\beta, g, a, k\}$ is the private-key.

(2) Encryption To encrypt n-bit binary plaintext $\mathbf{x} = (x_1, \ldots, x_n)$, we compute:

$$c = (c_1, c_2) = \prod_{i=1}^{n} (s_i, u_i)^{x_i} \mod p \tag{4}$$

and send ciphertext c to the receiver.

(3) Decryption To recover plaintext $\mathbf{x}$ from ciphertext c, the receiver should do the following:

(a) Compute

$$d = c_2.(c_1^{-1})^a \mod p = \frac{\prod_{i=1}^{n}(u_i)^{x_i}}{\prod_{i=1}^{n}(s_i^a)^{x_i}} \mod p = \prod_{i=1}^{n} p_i^{x_i} \mod p.$$

(b) Since $p > \prod_{i=1}^{n} p_i$ and $x_i \in \{0, 1\}$ hence $\prod_{i=1}^{n} p_i^{x_i} \mod p = \prod_{i=1}^{n} p_i^{x_i}$ and so we have

$$d = \prod_{i=1}^{n} p_i^{x_i}.$$

Since $x_i \in \{0, 1\}$, then d is the product of some distinct primes p_i. By Lemma 1, we conclude that

$$x_i = \begin{cases} 1 & \text{if} \quad p_i \mid d \\ 0 & \text{if} \quad p_i \nmid d. \end{cases} \qquad 1 \leq i \leq n$$

Security analysis

In the modified cryptosystem, we have

$$c_1 = \prod_{i=1}^{n} s_i^{x_i} \mod p = t^{\sum_{i=1}^{n} x_i} \mod p = t^h \mod p,$$

where $h = \sum_{i=1}^{n} x_i$ and $t = s_i = g^k \mod p$ are integers. Since discrete logarithm problem is intractable, so, we cannot determine Hamming weight h from $c_1 = t^h$

mod p and thus, the proposed attack is not feasible in this case.

Birthday Attack[7]

If the prime p is chosen too small, then from inequality $p > \prod_{i=0}^{n} p_i$, it follows that n is small. Hence p must be sufficiently large (we recommend at least $n \geq 1180$) to prevent birthday-search through two lists A and B of $2^{n/2}$ elements to find a couple of sets such that:

$$\prod_{i \in A} u_i = (\prod_{i \in B} u_i)^{-1}.c_2 \mod p.$$

6 Conclusion

In this paper, we considered a hybrid public key cryptosystem. This cryptosystem uses the ElGamal cryptosystem in the key generation stage for disguising the secure knapsack (private-key) in order to produce the public knapsack (public-key), and subset product (multiplicative knapsack) problem for encryption and decryption. We show that this combination leaks the security and makes the cryptosystem vulnerable to ciphertext-only attack. To avoid this attack, we compute the ciphertext modulo a large prime p. Moreover, we showed that the proposed cryptosystem is impractical. We modified this cryptosystem for increasing security and efficiency. In this case, if one wishes to break the cryptosystem, he/she must computes discrete logarithm problem which is infeasible.

References

1. S. Akleylek, L. Emmungil and U. Nuriyev, Algorithm for peer-to-peer security, *journal of Appl. Comput. Math.*, Vol. 6 (22), pp.258-264, 2007. Available in: `http://www.science.az/acm/v_6_n_2_2007/258-264.pdf`
2. T. ElGamal, A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithm, *IEEE Trans. on Information Theory*, Vol. 31(4), pp.469-472, 1985.
3. M.K. Lai, Knapsack Cryptosystems: The past and the future, Available in: `http://www.ics.uci.edu/~mingl/knapsack.html`.
4. R.C. Merkle and M.E. Hellamn, Hiding Information and Signatures in Trapdoor knapsacks, *IEEE Trans. on Information Theory*, Vol. 24, pp.525-530, 1978.
5. A. Menezes, P. van Oorschot and S. Vanstone, Handbook of Applied Cryptography. CRC Press, 1997.
6. B.M. Moret, "The Theory of Communication", Addison-Wesley, Reading 1978.
7. D. Naccache and J. Stern, A new public-key cryptosystem, In *EuroCrypt'1997*, LNCS, Vol. 1233, pp. , 1997.