Analysis of Privacy-Preserving Element Reduction of Multiset

Jae Hong Seo [1], HyoJin Yoon [2], Seongan Lim [3], Jung Hee Cheon [4] and Dowon Hong [5]

[1,4] Department of Mathematical Sciences and ISaC-RIM, Seoul National University, Seoul, 151-747, Korea.
(jhsbhs, jhcheon)@math.snu.ac.kr

[2] School of Electrical Engineering and Computer Sciences, Seoul National University, Seoul, 151-747, Korea.
jin25@math.snu.ac.kr

[3] Department of Mathematics, INHA University, Incheon, 402-751, Korea
seongannym@inha.ac.kr

[5] Electronics and Telecommunications Research Institute, Taejon, 305-700, Korea
dwhong@etri.re.kr

Abstract. Among private set operations, the privacy preserving element reduction of a multiset can be an important tool for privacy enhancing technology as itself or in the combination with other private set operations. Recently, a protocol, over-threshold-set-union-protocol, for a privacy preserving element reduction method of a multiset was proposed by Kissner and Song in Crypto 2005. In this paper, we point out that there is a mathematical flaw in their polynomial representation of element reduction of a multiset and the resulting protocol error from the flaw in the polynomial representation of a multiset. We correct their polynomial representation of a multiset and propose an over-threshold-set-operation-protocol based on the corrected representation. Our over-threshold-set-operation-protocol can be combined with a privacy preserving set operation and outputs those elements appears over the predetermined threshold number times in the resulting multiset of set operation.

Key words : Privacy-Preserving Operations, Set Operations, Element Reduction, Multi-party Computations

1 Introduction

Private set operations, such as *set intersection, set union* and *element reduction* of multisets, are important tools for privacy preserving of many applications. The *element reduction*(by d) method for a multiset S, a set allows a repetition of elements, is a method to get a multiset, denoted by $Rd_d(S)$, after reducing the repetition number of each element by d from the multiset S. Whenever one sees an element a in $Rd_d(S)$, then he/she knows that a appears more than d times in S. A private *element reduction* method of a multiset enables

controlled disclosure of private information and it can be combined with other private set operations to support controlled privacy level of the output of the private set operation. The *element reduction* of a multiset also can be used to develop privacy preserving techniques for distributed network monitoring. In distributed network monitoring service, each node monitors anomalous local traffic, and a distributed nodes collectively identify behaviors that are identified by at least a threshold t number of monitors.

Monitoring individuals' network usage requires appropriate privacy preserving of network users. In some cases, monitoring of individuals' network behaviors generates records subject to a system of protections under Privacy Act (e.g, FERPA). The privacy framework includes notification of policies; minimization of collection of data; limits on secondary use; nondisclosure and consent; a need to know before granting third parties access to data. On the other hand, in order to run normal network flow, it is necessary to control anomalous behaviors which could threat the normal flow of the network system. By using private element reduction method with respect to a threshold t on a multiset, one can identify just those elements appeared at least t times in the multiset but cannot obtain any information of the elements appeared less than t times. Hence the private element reduction supports an appropriate privacy preserving in monitoring network system within Privacy Act.

In Crypto 2005, Kissner and Song studied privacy-preserving set operations, such as set intersection, set union and element reduction of multisets ([6]). By using the polynomial representation of set operations and public key encryption scheme with homomorphic property, they proposed protocols for privacy preserving set intersection and set union. They also proposed a protocol, Over-Threshold Set-Union protocol, by using their polynomial representation of element reduction of a multiset.

In this paper, we point out that there is a mathematical flaw in their polynomial representation of element reduction of a multiset. Due to this mathematical flaw, it may happen to identify elements appeared in the multiset less than the threshold t as elements appeared at least t times in the multiset in their proposed Over-Threshold Set-Union protocol. Hence when we apply Over-Threshold Set-Union protocol to network monitoring system, it may happen to identify user with normal behavior as an user with an anomalous behavior and this leads privacy threat to normal user. Hence it could be a serious problem in privacy preserving techniques and it is necessary to correct the protocol. We give a correction in Kissner

and Song's polynomial representation of element reduction of a multiset. We also modify their Over-Threshold Set-Union protocol and propose an Over-Threshold Set-Operation protocol based on the corrected polynomial representation. Our over-threshold set-operation protocol can be combined with any privacy preserving set operation and outputs those elements appears in the resulting multiset of set operation over the predetermined threshold number. The security proof of Kissner and Song's protocol can be preserved in our protocol, since our protocol differs from Kissner and Song's protocol only in the way of representing the element reduction of a multiset as polynomials. Hence our corrected over-threshold set-operation protocol is provably secure in both standard adversary models: honest-but-curious(HBC) and malicious adversary.

Our paper is organized as follows. In Section 2, we introduce some preliminaries, such as adversary model and mathematical tools used in this paper. In Section 3, we describe Kissner and Song's polynomial representation of set union, set intersection and element reduction of a multiset and point out a mathematical flaw in the element reduction case and errors in their over-threshold set-union protocol. We correct Kissner and Song's polynomial representation of element reduction of a multiset and propose an over-threshold set-operation protocol based on the correction in Section 4. We conclude our paper in Section 5.

2 Preliminaries

In this section, we describe the adversary models and cryptographic tools used in this paper. Most can be found in [6, 7].

2.1 Adversary Model

In this paper, we propose a modified privacy-preserving element reduction protocol which is secure under the honest-but-curious adversary model, and it can be extended to a protocol which is secure under malicious adversary model. In our protocol, we consider standard adversary models : honest-but-curious adversary model and malicious adversary model, that can be described informally as follows [6, 7]. We can find the formal definitions of these models in [5].

Honest-But-Curious Adversary In this model, all parties act according to their prescribed actions in the protocol. Security in this model is straightforward: no player or

coalition of $c(< n)$ players (who cheat by sharing their private information) gains information about other players' private input sets, other than what can be deduced from the result of the protocol. This is formalized by considering an ideal implementation where a trusted third party (TTP) receives the inputs of the parties and outputs the result of the defined function. We require that in the real implementation of the protocol, that is, one without a TTP each party does not learn more information than in the ideal implementation.

Malicious Adversary In this model, an adversary may behave arbitrarily. In particular, we cannot hope to prevent malicious parties from refusing to participate in the protocol, choosing arbitrary values for its private input set, or aborting the protocol prematurely. Instead, we focus on the standard security definition (see, e.g., [5]) which captures the correctness and the privacy issues of the protocol. Informally, the security definition is based on a comparison between the ideal model and a TTP, where a malicious party may give arbitrary input to the TTP. The security definition is also limited to the case where at least one of the parties is honest. Let Γ be the set of colluding malicious parties; for any strategy Γ can follow in the real protocol, there is a translated strategy that it could follow in the ideal model, such that, to Γ, the real execution is computationally indistinguishable from execution in the ideal model.

2.2 Homomorphic Public key Encryption scheme

To perform the privacy-preserving set operation without trusted third party, we need a public key encryption scheme with special feature. The requirements of the encryption scheme are as follows:

1. The encryption scheme should be additively homomorphic. That is,
   - For given $E(a)$ and $E(b)$, $E(a+b) := E(a) +_h E(b)$ can be computed efficiently, where '$+_h$' is an additive operation in the image of the encryption function E.
   - For a constant c and $E(a)$, $E(c \cdot a) := c \times_h E(a)$ can be computed efficiently, where '$\times_h$' is an multiplicative operation in the image of the encryption function E.,

   where $E(\cdot)$ is an encryption function of an additively homomorphic encryption scheme.

2. The ciphertext should be re-randomized.

3. The encryption scheme should support (n, n)-threshold decryption, i.e. the corresponding private key is shared by a group of n players and decryption must be performed by all n players acting together.

The Paillier's cryptosystem [9] satisfies these requirements.

2.3 Multiset and its Polynomial Representation

We consider the concept of *multiset*. Differently from an ordinary "set", a multiset permits duplication of its elements. For example, in a multiset, an element is represented more than once like $\{a, a, b\}$ and the multiset is different from $\{a, b\}$.

Now, we define set intersection, and set union of multisets and the element reduction of a multiset as follows:

Definition 1 *The union of multisets A and B, $A \cup B$, is the multiset composed of the elements which are in A or B. If an element 'a' appears l_A times in A and l_B times in B, then 'a' appears $l_A + l_B$ times in $A \cup B$.*

Definition 2 *The intersection of multisets A and B, $A \cap B$, is the multiset composed of the elements which are in both A and B. If an element 'a' appears l_A times in A and l_B times in B, then 'a' appears $min\{l_A, l_B\}$ times in $A \cap B$.*

Definition 3 *The element reduction by d, $Rd_d(A)$, of a multiset A is the multiset composed of the elements of A such that for every element 'a' that appears d' times in A, 'a' is included $max\{0, d' - d\}$ times in $Rd_d(A)$.*

Polynomial Representation of a Multiset Kissner and Song use a homomorphic encryption scheme to achieve the property of privacy-preserving in set operation. Let a ring R be the domain of the homomorphic encryption function and P a subset of the ring R, where the elements in P are uniformly distributed in R and the probability that randomly chosen element of R is an element in P is negligible.

Kissner and Song consider the multiset S whose elements belong to P and define the polynomial representation of the multiset as follows:

– *From a multiset S to a polynomial $f_S \in R[x]$:*

- Given a multiset $S = \{S_j\}_{1 \le j \le k}, S_j \in P$, the polynomial $f_S \in R[x]$ represents the multiset S can be constructed as

$$f_S(x) = \prod_{1 \le j \le k} (x - S_j).$$

– *From a polynomial $f \in R[x]$ to a multiset S:*

- Given a polynomial $f \in R[x]$, the multiset S represented by f can be defined as follows:

$$a \in S \text{ and } a \text{ appears } t \text{ times in } S \iff$$
$$(x-a)^t | f,\ (x-a)^{t+1} \not| f \text{ and } a \text{ is an element in } P$$

Feasible Homomorphic Operations of Encrypted Polynomials We define an encryption $E(f(x))$ of a polynomial $f(x) = \sum_{i=0}^{\deg(f)} f[i]x^i$ as a tuple of each encryption of coefficient $f[i]$, where $0 \le i \le \deg(f)$. That is,

$$E(f(x)) := (E(f[0]), \cdots, E(f[\deg(f)])).$$

If the public key encryption scheme satisfies the additive homomorphic property, we can perform the following computations without its decryption key.

1. The addition of encrypted polynomials: Given $E(f_1)$ and $E(f_2)$, we can compute $E(f_1+f_2)$ as follows:

$$E((f_1 + f_2)[i]) := E(f_1[i]) +_h E(f_2[i])\ (0 \le i \le max\{\deg(f_1), \deg(f_2)\}).$$

2. The addition of encrypted polynomial and unencrypted polynomial: Given $E(f_1)$ and f_2, we can compute $E(f_1 * f_2)$ as follows:

$$E((f_1 * f_2)[i]) := (f_2[0] \times_h E(f_1[i])) +_h \cdots +_h (f_2[i] \times_h E(f_1[0]))$$
$$(0 \le i \le \max\{\deg(f_1) + \deg(f_2)\}).$$

3. The differentiation of encrypted polynomial: Given $E(f)$, we can compute $E(\frac{df}{dx})$ as follows:

$$E\left(\frac{df}{dx}[i]\right) := (i+1) \times_h E(f_1[i+1])\ (0 \le i \le \deg(f_1) - 1).$$

4. The value of encrypted polynomial at plain point: Given $a \in R$ and $E(f)$, we can compute $E(f(a))$ as follows:

$$E(f(a)) := (a^0 \times_h E(f[0])) +_h \cdots +_h (a^{\deg(f)} \times_h E(f[\deg(f)])).$$

3 Analysis of Kissner and Song's Element Reduction Methods

In this section, we review the results of Kissner and Song [6, 7] and point out a flaw in their polynomial representation of element reduction of a multiset. Furthermore we show that their over-threshold set-union protocol has critical errors since they used the incorrect polynomial representation of element reduction of a multiset.

3.1 Errors of Kissner and Song's Polynomial Representation for Element Reduction

Kissner and Song use polynomials to represent multisets and propose probabilistic polynomial representations corresponding to set union, set intersection, and element reduction of a multiset. Kissner and Song's polynomial representations of the set union, set intersection and *incorrect* element reduction (by d) of a multiset are given as follows:

Union Let f and g be polynomial representations of multisets S and T, respectively. The polynomial $f * g$ is the polynomial representation of multiset $S \cup T$.

Intersection Let f and g be polynomial representations of multisets S and T, respectively. For random polynomials r, s of higher to or same degree with $\deg(f)$, $f*r+g*s$ is equal to $\gcd(f,g)*u$, where u is a uniformly distributed in $R^{\alpha}[x]$, $R^{\alpha}[x]$ is the set of all polynomials whose coefficients are in R and degrees are lower to or same with $\alpha = 2\deg(f) - |S \cap T|$. The polynomial $f*r+g*s$ is a polynomial representation of multiset $S \cap T$ with overwhelming probability.

(Incorrect) Element Reduction (by d) Let f be the polynomial representation of a multiset S. For random polynomials r, s of degree $\deg(f)$ or more and a random polynomial F with degree d whose solutions are not in P, $f^{(d)} * F * r + f * s$ is equal to $\gcd(f, f^{(d)}) * u$, where $f^{(d)}$ is the d-th derivative of f, u is a uniformly distributed in $R^{\alpha}[x]$, $R^{\alpha}[x]$ is the set of all polynomials whose coefficients are in R and degrees are lower to or same with $\alpha = 2\deg(f) - |Rd_d(S)|$. The polynomial $f^{(d)} * F * r + f * s$ is a polynomial representation of multiset $Rd_d(S)$ with overwhelming probability.

In the above, since u is uniformly distributed in $R^{\alpha}[x]$, the probability that u has a root in P is negligible. Thus $f * r + g * s = \gcd(f, g) * u$ is the polynomial representation of the

multiset $S \cap T$ and $f^{(d)} * F * r + f * s = \gcd(f, f^{(d)}) * u$ is the polynomial representation of the multiset $Rd_d(S)$ with overwhelming probability. The polynomial representation of element reduction proposed in [6, 7] uses following lemma.

Lemma 1 (Lemma 2 in [6]) *Let R be a ring and $f(x) \in R[x]$. For $(d \geq 1)$,*

(1) *if $(x-a)^{d+1} | f(x)$, then $(x-a) | f^{(d)}(x)$.*
(2) *if $(x-a) | f(x)$ and $(x-a)^{d+1} \not| f(x)$, then $(x-a) \not| f^{(d)}(x)$.*

By using Lemma 1, Kissner and Song showed that $\gcd(f, f^{(d)})$ is a polynomial representation of $Rd_d(S)$, where f is the polynomial representation of the multiset S. But, in general, Lemma 1 is incorrect when $d > 1$. We can make a counter-example of Lemma 1 as follows.

Example 1 *Let a,b and c be distinct elements of ring R. Let $f(x) = (x-a)(x-b)(x-c)$. If the Lemma 1 is correct then the following relation:*

$$(x-a) \not| f^{(2)}(x)$$

holds since $(x-a) | f$ and $(x-a)^3 \not| f(x)$. But $f^{(2)}(x) = 6x - 2(a+b+c)$ and $(x - \frac{(a+b+c)}{3}) | f^{(2)}(x)$ i.e.

$$(x-a) | f^{(2)}(x), \text{ when } c = 2a - b.$$

This contradicts Lemma 1.

Thus we see that Lemma 1 is incorrect. The mathematical flaw in Lemma 1 results errors in their polynomial representation of element reduction. In the same example as in Example 1, we consider the element reduction by 2 for the set $S = \{a, b, c\}$ of distinct elements with $c = 2a - b$. Clearly, we see that $Rd_2(S) = \phi$. But as we have in the above example, $\gcd(f, f^{(2)}) = (x-a)$, which cannot be a polynomial representation of $Rd_2(S) = \phi$.

3.2 Analysis of the Kissner and Song's Protocol

Now, we analyze the *Over-Threshold Set-Union protocol* proposed in [6, 7]. The *Over-Threshold Set-Union protocol* is a multiparty protocol with n users under the assumption that at most $c (< n)$ players can dishonestly collude. A user i (where $1 \leq i \leq n$) generates a multiset S_i

whose elements represent private information and they are in P. Assume that each individual multiset should have the same cardinality. That is, for all i such $1 \leq i \leq n$, $|S_i| = k$ for some k. The j-th element of a multiset S_i is represented by $(S_i)_j$. At the end of the protocol, all users wants to get a multiset which consists of the over-threshold elements in the set union $S = S_1 \cup \cdots \cup S_n$ of each user's multiset. The goal of the protocol is to solve the *Over-Threshold Set-Union problem* which is defined in [6, 7] as follows:

Definition 4 [1] *All players know elements in the union of the each players' private multisets that appears more than a threshold number d times, and the frequency of these elements in the union without gaining any other information. We call the elements of resulting set as over-threshold elements in the union of private sets of all players.*

In their *Over-Threshold Set-Union protocol*, Kissner and Song use the element reduction method to obtain over-threshold elements in set union. Let a fixed threshold number be t and a polynomial p be the corresponding to the multiset S of union of private sets of all players. As shown in the previous section, Kissner and Song computed $\gcd(p, p^{(t-1)})$ as the polynomial representation of $Rd_{t-1}(S)$ thus $\gcd(p, p^{(t-1)})$ doesn't give the correct representation of $Rd_{t-1}(S)$ in some cases.

Now, we show that their protocol outputs wrong results in the case of Example 1. We apply their *Over-Threshold Set-Union protocol* to the set union $S = \{a, b, 2a - b\}$ with the threshold 3. Then, the protocol outputs the corresponding set $\{a\}$ as the set of the over-threshold 3 in set union. But a appears only once in the set S, hence Kissner and Song's protocol is not a correct threshold 3 protocol.

Suppose we consider the above example in the distributed network monitoring system with a privacy policy that says *'the monitoring system identify only the users with anomalous behavior over threshold* 3*'*. Then the user 'a' will be identified in the monitoring system, but it appears only once and should not be identified in the monitoring system. This conflicts the privacy policy they adopt. Hence a correction in Kissner and Song's polynomial representation of element reduction is required.

[1] This definition can be extended to a Over-Threshold Set-Operation problem by exchanging the union for general set operation.

4 A Correct Polynomial Representation and an Over-Threshold Set-Operation Protocol

In this section, we suggest a correct polynomial representation of element reduction and propose a new over-threshold set-operation protocol using the corrected polynomial representation.

4.1 A Correct Polynomial Representation of Element Reduction

We propose new method of element reduction by correcting Lemma 1. Particularly, we prove $\gcd(f, f', \cdots, f^{(d)})$ is a polynomial representation of $Rd_d(S)$. By correcting Lemma 1, we have following lemma.

Lemma 2 *Let $f(x) \in R[x]$. The followings are equivalent.*

(1) $(x-a)^{d+1} \mid f(x)$.

(2) $f(a) = f'(a) = \cdots = f^{(d)}(a) = 0$, *i.e.* $(x-a) \mid f,\ (x-a) \mid f',\ \cdots,\ (x-a) \mid f^{(d)}$.

Proof. ((1) $\Rightarrow$ (2)) Suppose $(x-a)^{d+1} \mid f(x)$. Then $f(x) = (x-a)^{d+1}g(x)$ for some $g(x)$, and clearly we have $f(a) = 0$. We also have $f'(x) = (d+1)(x-a)^d g(x) + (x-a)^{d+1}g'(x)$ and it gives $f'(a) = 0$. For $1 \leq n \leq d$, we get $f^{(n)}(x) = (d+1)\cdots(d-n+2)(x-a)^{d-n+1}g(x) + (x-a)^{d-n+2}h_n(x)$ for some $h_n(x)$. Hence $f(a) = \cdots = f^{(d)}(a) = 0$.

((2) $\Rightarrow$ (1)) First we will show that if $f^{(n)}(a) = 0$ and $(x-a)^n | f(x)$ then $(x-a)^{n+1} \mid f(x)$. Since $(x-a)^n | f(x)$, we have $f(x) = (x-a)^n g(x)$ for some $g(x)$. $f^{(n)}(x) = n!g(x) + (x-a)h_n(x)$ for some $h_n(x)$. Since $f^{(n)}(a) = 0$, we have $g(a) = 0$, which implies that $g(x) = (x-a)g_1(x)$ for some $g_1(x)$. Therefore $f(x) = (x-a)^{n+1}g_1(x)$.
Because $f^{(1)}(a) = 0$ and $(x-a)|f(x)$ by hypothesis, we have $(x-a)^2|f(x)$. And again together $(x-a)^2|f(x)$ with $f^{(2)}(a) = 0$, we have $(x-a)^3|f(x)$. By repeating the same procedure with $f^{(3)}(a) = 0, \cdots, f^{(d)}(a) = 0$, eventually we get $(x-a)^{d+1}|f(x)$. □

We obtain the following Corollary from the Lemma 2.

Corollary 3 $(x-a)^{d+1} \mid f(x) \Rightarrow (x-a)^d \mid f'(x)$.

Proof. By Lemma 2,

$$\begin{aligned}(x-a)^{d+1} \mid f(x) &\Leftrightarrow f(a) = f'(a) = \cdots = f^{(d)}(a) = 0 \\ &\Rightarrow f'(a) = f''(a) = \cdots = f'^{(d-1)}(a) = 0 \\ &\Rightarrow (x-a)^{d} \mid f'(x) \text{ by applying Lemma 2 to } f'.\end{aligned}$$

Therefore Corollary 3 is proved. □

Now, we will prove $\gcd(f, f', \cdots, f^{(d)})$ is a correct polynomial representation of $Rd_d(S)$.

Theorem 4 *Let a polynomial f be a polynomial representation of a multiset S. For $a \in S$ and positive integer ℓ_a,*

$$\begin{aligned}(x-a)^{\ell_a} \mid \gcd(f, f', \cdots, f^{(d)}), \quad (x-a)^{\ell_a+1} \nmid \gcd(f, f', \cdots, f^{(d)}) \\ \Leftrightarrow \ a \text{ appears } \ell_a \text{ times in } Rd_d(S).\end{aligned}$$

i.e. $\gcd(f, f', \cdots, f^{(d)})$ *is a polynomial representation of* $Rd_d(S)$.

Proof. ($\Rightarrow$) Suppose that ℓ_a is a positive integer satisfying

$$(x-a)^{\ell_a} \mid \gcd(f, \cdots, f^{(d)}), \text{ and } (x-a)^{\ell_a+1} \nmid \gcd(f, \cdots, f^{(d)}).$$

Then since $(x-a)^{\ell_a} \mid \gcd(f, \cdots, f^{(d)})$, we have

$$(x-a)^{\ell_a} \mid f, \ (x-a)^{\ell_a} \mid f', \ \cdots, \ (x-a)^{\ell_a} \mid f^{(d)}.$$

By Corollary 3, we have

$$(x-a)^{\ell_a-1} \mid f^{(d+1)}, \ \cdots, \ (x-a)^{1} \mid f^{(d+\ell_a-1)}.$$

Thus we have $(x-a)^{d+\ell_a} \mid f$ by Lemma 2.

If $(x-a)^{\ell_a+d+1} | f$, then $f(a) = \cdots = f^{(\ell_a+d)}(a) = 0$. The part $f(a) = \cdots = f^{(\ell_a)}(a) = 0$ implies that $(x-a)^{\ell_a+1} | f$ by Lemma 2. Similarly, $f^{(i)}(a) = \cdots = f^{(\ell_a+i)}(a) = 0$ implies that $(x-a)^{\ell_a+1} | f^{(i)}$ for all i with $1 \le i \le d$ and it means that $(x-a)^{\ell_a+1} | \gcd(f, \cdots, f^{(d)})$. This contradicts to the hypothesis. Therefore, we have $(x-a)^{\ell_a+d+1} \nmid f$.

Hence ℓ_a satisfies

$$(x-a)^{\ell_a+d} | f \text{ and } (x-a)^{\ell_a+d+1} \nmid f.$$

And we know that a appears $\ell_a + d$ times in multiset S since f is a polynomial representation of S. In conclusion, a appears ℓ_a times in $Rd_d(S)$.

($\Leftarrow$) Suppose that a appears ℓ_a times in $Rd_d(S)$, then we have $(x-a)^{\ell_a+d} \mid f$ and $(x-a)^{\ell_a+d+1} \not| f$. By Corollary 3, we have

$$(x-a)^{\ell_a+d-1} \mid f',\ \cdots,\ (x-a)^{\ell_a} \mid f^{(d)}, \cdots,\ (x-a) \mid f^{(\ell_a+d-1)},\ (x-a) \not| f^{(\ell_a+d)},$$

which implies that

$$(x-a)^{\ell_a} \mid f,\ (x-a)^{\ell_a} \mid f', \cdots,\ (x-a)^{\ell_a} \mid f^{(d)},\ \text{but}\ (x-a)^{\ell_a+1} \not| f^{(d)}.$$

Thus ℓ_a is a positive integer satisfying

$$(x-a)^{\ell_a} |\ \gcd(f, \cdots, f^{(d)}) \text{ and } (x-a)^{\ell_a+1} \not| \gcd(f, \cdots, f^{(d)}). \quad \square$$

By Theorem 4, we can correct the Kissner and Song's polynomial representation of element reduction as follows:

Correct Element Reduction (by d) Let f be the polynomial representation of a multiset S. For random polynomial r_i's of degree $\deg(f)$ or more and random polynomial F_i's with degree i whose solutions are not in P, $\sum_{i=0}^{d} f^{(i)} * F_i * r_i$ is equal to $\gcd(f, f', \ldots, f^{(d)}) * u(x))$. Thus the polynomial $\sum_{i=0}^{d} f^{(i)} * F_i * r_i$ is a polynomial representation of multiset $Rd_d(S)$ with overwhelming probability.

4.2 An Over-Threshold Set-Operation Protocol

Now, we propose an Over-Threshold Set-Operation protocol using our correct polynomial representation of element reduction of a multiset. As described above, the goal of this protocol is for all players to obtain the multiset of elements which appear in the result of set operation of each private multiset more than a predetermined threshold number without gaining any other information.

There are $n(\geq 2)$ honest-but-curious players with a private input set S_i such that $|S_i| = k$. We assume that at most $c(< n)$ players can dishonestly collude. The players share the secret

key sk corresponding to the public key pk for a homomorphic cryptosystem which supports threshold group decryption. Let the threshold number be d and F_j be an arbitrary polynomial of degree j which has no roots representing elements of the set P.

Protocol : Over–Threshold Set–Operation–HBC

1. **Set-Operation** Each player $i = 1, \cdots, n$ computes $f_i(x) = (x - (S_i)_1) \cdots (x - (S_i)_k)$. Players perform the predetermined set operation protocol and player 1 obtain the encryption of polynomial p, corresponding to the result of set operation, $E_{pk}(p)$. Player 1 distributes $E_{pk}(p)$ to players $2, \cdots, c+1$.
2. **Element-Reduction** Each player $i = 1, \cdots, c+1$
   - **(a)** Computes $E_{pk}(p'), \cdots, E_{pk}(p^{(d)})$ from $E_{pk}(p)$.
   - **(b)** Chooses randomly $d+1$ polynomials $t_{i,0}, \cdots, t_{i,d} \in R^k[x]$.
   - **(c)** Sends $E_{pk}(p * t_{i,0} + F_1 * p' * t_{i,1} + \cdots + F_d * p^{(d)} * t_{i,d})$ to all other player.
3. **Group-Decryption** All players perform a group decryption to obtain $\Phi = F_d * p^{(d)} * (\sum_{i=1}^{c+1} t_{i,d}) + \cdots + F_1 * p' * (\sum_{i=1}^{c+1} t_{i,1}) + p * (\sum_{i=1}^{c+1} t_{i,0})$.
4. **Recovering-Set** Each player $i = 1, \cdots, n$ determines the resulting set depending on a kind of set operation.

We use the modified element reduction method of the multiset in the step 2 and fix the flaw of the original one proposed by Kissner and Song. The step 1 and 4 of the above protocol can be varied according to a kind of set operation. In [6, 7], protocols for privacy preserving set operations, set union and set intersection, were proposed. In step 1, if we apply Kissner and Song's privacy-preserving (set union/set intersection) then we obtain (Over-Threshold Set-Union protocol /Over-Threshold Set-Intersection protocol), respectively.

Because the difference of Over–Threshold Set–Union protocol proposed in [6, 7] from our protocol is only the polynomial representation method of element reduction of multiset, it does not affect the security of the protocol. Thus, the our protocol has the same security with that of [6, 7] in the set intersection and set union cases when we follow their set operation protocol.

5 Conclusion

A privacy preserving element reduction method can be an important tool to identify badly behaved internet users while it preserves privacy for normal users. In Crypto 2005, Kissner and Song introduced a method of polynomial representation of element reduction of a multiset and proposed an Over-Threshold-Set-Union protocol using the polynomial representation and homomorphic public key encryption scheme. In this paper, we have shown that their polynomial representation is not correct and its impact to their protocol can be somewhat critical for privacy preserving techniques. We present a correction for the polynomial representation of element reduction of a multiset in this paper. We also modified their Over-Threshold Set-Union protocol and proposed an Over-Threshold Set Operation protocol based on the corrected polynomial representation. Our Over-threshold-set-operation-protocol can be combined with a privacy preserving set operation and outputs those elements appears over the predetermined threshold number times in the resulting multiset of set operation.

References

1. D. Boneh, E-J. Goh, and K. Nissim. Evaluating 2-DNF formulas on ciphertexts. In *TCC2005*, LNCS Vol. 3378, pp. 325–341. Springer-Verlag, 2005.
2. R. Agrawal, A. Evfimievski, and R. Srikant. Information sharing across private databases. In *SIGMOD 2003*, Proceedings of the 2003 ACM SIGMOD international conference on Management of data, pp. 86–97, ACM Press, 2003.
3. M. Freedman, K. Nissim, and B. Pinkas. Efficient private matching and set intersection. In *Advances in Cryptology - Eurocrypt 2004*, LNCS Vol. 3027, pp. 1–19, Springer-Verlag, 2004.
4. P-A. Fouque and D. Pointcheval. Threshold cryptosystems secure against chosenciphertext attacks. In *Advances in Cryptology - Asiacrypt 2000*, LNCS Vol. 1976, pp. 573–584, Springer-Verlag, 2000.
5. O. Goldreich, The Foundations of Cryptography - Vol. 2, Cambridge University Press, 2004.
6. L. Kissner and D. Song. Privacy-preserving set operations. In *Advances in Cryptology - Crypto 2005*, LNCS Vol. 3621, pp. 241–257, Springer-Verlag, 2005.
7. L. Kissner and D. Song. Private and threshold set-intersection. Technical Report CMU-CS-05-113, Carnegie Mellon University, 2005
8. T. Okamoto and S. Uchiyama. A new public-key cryptosystem as secure as factoring. In *Advances in Cryptology - Eurocrypt 1998*, LNCS Vol. 1403, pp. 308–318. Springer-Verlag, 1998.
9. P. Pallier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology - Eurocrypt 1999*, LNCS Vol. 1592 , pp. 223–238, Springer-Verlag, 1999.