Efficient Oblivious Transfer in the Bounded-Storage Model

Dowon Hong, Ku-Young Chang, and Heuisu Ryu

Information Security Research Division, ETRI
161 Gajeong-Dong, Yuseong-Gu, Daejeon, 305-350, KOREA
{dwhong,jang1090,hsryu}@etri.re.kr

Abstract. In this paper we propose an efficient OT_1^N scheme in the bounded storage model, which is provably secure without complexity assumptions. Under the assumption that a public random string of M bits is broadcasted, the protocol is secure against any computationally unbounded dishonest receiver who can store τM bits, $\tau < 1$. The protocol requires the sender and the receiver to store $N \cdot O(\sqrt{kM})$ bits, where k is a security parameter. When $N = 2$, our protocol is similar to that of Ding [10] but has more efficient round and communication complexities. Moreover, in case of $N > 2$, if the sender and receiver can store $N \cdot O(\sqrt{kM})$ bits, we are able to construct a protocol for OT_1^N which has almost the same complexity as in OT_1^2 scheme. Ding's protocol was constructed by using the interactive hashing protocol which is introduced by Noar, Ostrovsky, Venkatesan and Yung [15] with very large round-complexity. We propose an efficiently extended interactive hashing and analyze its security. This protocol answers partially an open problem raised in [10].

1 Introduction

Consider two parties of the sender Alice and the receiver Bob. Alice has N secret bits $X_0, X_1, \cdots, X_{N-1} \in GF(2)$, and Bob has a secret value $c \in \{0, 1, \cdots, N-1\}$. Alice sends $X_0, X_1, \cdots, X_{N-1}$ in such a way that Bob receives X_c, but does not learn any information about other secrets X_i, $i \neq c$, and Alice learns nothing about c. An 1-out-of-N Oblivious Transfer (OT_1^N) is a cryptographic two-party protocol that provides a solution for the goal.

OT_1^2 was suggested by Even, Goldreich, and Lempel [11], as a generalization of Rabin's Oblivious Transfer (OT) [16], and Crépeau [6] proved that OT and OT_1^2 are equivalent. OT_1^N was introduced by Brassard, Crépeau, and Robert [2] under the name ANDOS (all or nothing disclosure of secrets). Oblivious transfer can be used to construct cryptographic protocols, such as bit commitment, zero-knowledge proof, and generally secure multi-party computation [13, 21, 12, 7, 14].

Traditionally, oblivious transfer has been constructed under complexity assumptions, such as the hardness of factoring or discrete log, or the existence of trapdoor one-way permutations. However, they do not guarantee information-theoretic security, and the security of the protocol could be subverted later, when enabled by breakthroughs in computing technology and algorithms. For example, protocols based on the hardness of factoring or computing discrete logarithms will become insecure if quantum computers become available [18]. Alternatives to computational security assumptions that have been proposed include quantum cryptography, the noisy channel model, and the bounded-storage model [1, 8, 3].

Cachin, Crépeau, and Marcil [4] proposed the first protocol for OT_1^2 in the bounded-storage model that is unconditionally secure, without any complexity assumption. Under the assumption that a public random string of M bits is broadcasted, the CCM protocol [4] guarantees provable security against any computationally unbounded dishonest receiver who can store τM bits, $\tau < 1$. Furthermore, the security against a dishonest receiver is preserved regardless of future increases in storage capacity. The case where the storage bound is placed on the sender is equivalent by the reversibility of OT [9]. Protocols in the bounded-storage model make use of a very large amount of auxiliary information, called public random string [17], in order to defeat the adversary. The public random string could be a random bit sequence broadcasted by a satellite or transmitted between the legitimate parties, or the signal of a deep-space radio source. Recently, Ding [10] proposed a similar but more efficient protocol for OT_1^2 in the bounded-storage model than the CCM protocol. Ding's protocol reduced the storage requirement from $O(M^{2/3})$ in the CCM protocol, to $O(\sqrt{kM})$ where k is a security parameter and proved that any dishonest receiver who stores $O(M)$ bits succeeded with probability at most $2^{-O(k)}$, rather than inverse polynomially small.

In this paper, we propose a provably secure and efficient protocol for OT_1^N with a storage-bounded receiver, without any complexity assumption. Our protocol uses N public random strings of M bits and requires the sender and the receiver to store $N \cdot O(\sqrt{kM})$ bits, where k is a security parameter. When $N = 2$, our protocol is similar to that of Ding's protocol but has more efficient round and communication complexities. Moreover, in case of $N > 2$, if the sender and the receiver can store $N \cdot O(\sqrt{kM})$ bits, we are able to construct a protocol for OT_1^N which has almost the same complexity as in OT_1^2 scheme. This is constructed based on an *extended interactive hashing* scheme.

Noar, Ostrovsky, Venkatesan and Yung [15] introduced the interactive hashing protocol, and Cachin, Crépeau, and Marcil [4] gave a new elegant analysis on it. Interactive hashing is a protocol between a challenger Alice with no input and a responder Bob with input string χ and provides a way to isolate two strings. One of the strings is Bob's input χ and the other is chosen randomly, without influence from Bob. However, Alice does not learn that which one is χ. Up to the present, the interactive hashing has been based on NOVY protocol [15] which has very large round and communication complexities. The round and communication complexities of NOVY protocol, which has the string of t bits to be transmitted, are $t-1$ rounds and t^2-1 bits respectively. Thus Ding's protocol for OT_1^2 which is based on NOVY protocol has very large round and communication complexities.

We propose more efficiently extended interactive hashing scheme than the NOVY protocol. We can accomplish the interactive hashing with $t/m-1$ rounds and $t^2/m-m$ bits of communication complexity, when m is a divisor of t, and provide a way to isolate more than two strings. As a concrete example of what is claimed in this paper (Section 4), assume that the length of a public random string is one Petabit, (i.e. $M = 10^{15}$), and $1000 \leq k \leq 10000$ for a security parameter k, then we can choose k easily such that the protocol has $t^{3/2}-t^{1/2}$-bit communication complexity which is much lower than that of NOVY protocol. This result answers partially an open problem raised in [10].

This paper is organized as follows. In Section 2, we construct a new universal hash family. Using this, we propose an extended interactive hashing protocol. The protocol for OT_1^N in the bounded-storage model is presented in Section 3. In Section 4, we discuss the complexity of our protocol.

2 Extended Interactive Hashing

In this section we propose an efficiently extended interactive hashing protocol and give an analysis on it. In order to construct this, we first introduce a new universal hash family.

2.1 Universal hash family

The technique of universal hashing was introduced in 1979 by Carter and Wegman [5] and is used in many areas of computer science and cryptography [19, 20].

Definition 1 *Let $\mathcal{F}$ be the set of all functions from X to Y and let $\mathcal{H}$-hash family be a subset of $|\mathcal{H}|$ functions in $\mathcal{F}$. $\mathcal{H}$-hash family is called universal if, for any distinct elements $x_1, x_2 \in X$, there exist at most $|\mathcal{H}|/|Y|$ functions $h \in \mathcal{H}$ such that $h(x_1) = h(x_2)$.*

Let t and m be positive integers such that m is a divisor of t. We now define a universal hash family from $GF(2)^t$ to $GF(2)^m$. Let $f(x)$ be an irreducible polynomial of degree m over $GF(2)$. Then $GF(2^m) = GF(2)[x]/(f(x))$ is represented as $\{\sum_{i=0}^{m-1} a_i x^i \ : \ a_i \in GF(2)\}$. Define the bijective function $\phi \ : \ GF(2)^m \rightarrow GF(2^m)$ by $(a_{m-1}, \cdots, a_1, a_0) \mapsto a_{m-1}x^{m-1} + \cdots + a_1 x + a_0$. Let $t = lm$. Then $GF(2)^t = (GF(2)^m)^l = \{(A_{l-1}, \cdots, A_1, A_0) \ : \ A_i \in GF(2)^m, 0 \leq i \leq l-1\}$. We regard $GF(2)^t$ as $(GF(2)^m)^l$ and let $S = (GF(2)^m)^l$. In order to define a universal hash family from S to $GF(2)^m$, for any $\zeta = (\zeta_{l-1}, \cdots, \zeta_1, \zeta_0) \in S$, we define the hash function using the above function ϕ as follows ;

$$\begin{aligned} h_\zeta \ : \ S &\longrightarrow GF(2)^m \\ (A_{l-1}, \cdots, A_1, A_0) &\longmapsto \phi^{-1}\left(\sum_{i=0}^{l-1} \phi(A_i) \cdot \phi(\zeta_i)\right). \end{aligned} \tag{1}$$

Consider the set $\mathcal{H}$ of hash functions from S to $GF(2)^m$ as follows ;

$$\mathcal{H} = \{h_\zeta \ : \ \zeta = (\zeta_{l-1}, \cdots, \zeta_1, \zeta_0) \in S\},$$

where h_ζ is defined in (1).

Lemma 1. *$\mathcal{H}$ is a universal hash family.*

Proof. For any two distinct elements $x = (x_{l-1}, \cdots, x_0), y = (y_{l-1}, \cdots, y_0) \in S$, we need to count the number of $\zeta = (\zeta_{l-1}, \cdots, \zeta_0) \in S$ with $h_\zeta(x) = h_\zeta(y)$. Since $x \neq y$, there is an index $i_0 \in \{0, \cdots, l-1\}$ such that $x_{i_0} \neq y_{i_0} \in GF(2)^m$. Then for any $\zeta \in S$

$$\begin{aligned} h_\zeta(x) = h_\zeta(y) &\Leftrightarrow \phi(\zeta_{i_0})(\phi(y_{i_0}) - \phi(x_{i_0})) + \sum_{i \neq i_0} \phi(\zeta_i)(\phi(y_i) - \phi(x_i)) = 0 \\ &\Leftrightarrow \phi(\zeta_{i_0})(\phi(y_{i_0}) - \phi(x_{i_0})) = \sum_{i \neq i_0} \phi(\zeta_i)(\phi(y_i) - \phi(x_i)) \in GF(2^m) \\ &\Leftrightarrow \phi(\zeta_{i_0}) = \sum_{i \neq i_0} \phi(\zeta_i)(\phi(y_i) - \phi(x_i)) \cdot (\phi(y_{i_0}) - \phi(x_{i_0}))^{-1}. \qquad (2) \end{aligned}$$

Since ϕ is bijective, for each choice of ζ_i's for $i \neq i_0$, equation (2) has exactly one solution in ζ_{i_0}. Since the number of i's for $i \neq i_0$ is $l-1$ and $\zeta_i \in GF(2)^m$ for each i, there are exactly $2^{m(l-1)} = |\mathcal{H}|/2^m$ functions $h_\zeta \in \mathcal{H}$ with $h_\zeta(x) = h_\zeta(y)$. Thus, $\mathcal{H}$ is a universal hash family. □

The universal hash family $\mathcal{H}$ defined above has the following properties.

Lemma 2. *Let $\mathcal{H}$ be the hash family defined above. For any two nonzero distinct elements $x,\ y \in S$ and for any $b \in GF(2)^m$, let $T_b = \{h \in \mathcal{H} \ : \ h(x) = b,\ h(y) = b\}$. Then $|T_b| = |\mathcal{H}|/2^{2m}$.*

Proof. For any two nonzero elements $x = (x_{l-1}, \cdots, x_0)$, $y = (y_{l-1}, \cdots, y_0) \in S$, let $x \neq y$. Note that $T_b = \{\zeta \in S : h_\zeta(x) = b,\ h_\zeta(y) = b\}$ by definition. Since $x \neq y$, there are two distinct indices $j,\ k \in \{0, \cdots, l-1\}$ such that $x_j \neq 0,\ y_k \neq 0 \in GF(2)^m$. Then for any $\zeta = (\zeta_{l-1}, \cdots, \zeta_0) \in S$

$$\begin{aligned} h_\zeta(x) = b &\Leftrightarrow \phi(x_j)\phi(\zeta_j) + \sum_{i \neq j} \phi(x_i)\phi(\zeta_i) = \phi(b) \\ &\Leftrightarrow \phi(\zeta_j) = (\sum_{i \neq j} \phi(x_i)\phi(\zeta_i) + \phi(b)) \cdot \phi(x_j)^{-1}, \\ h_\zeta(y) = b &\Leftrightarrow \phi(\zeta_k) = (\sum_{i \neq k} \phi(y_i)\phi(\zeta_i) + \phi(b)) \cdot \phi(y_k)^{-1}. \end{aligned}$$

Hence by similar method in Lemma 1, $|T_b| = 2^{m(l-2)} = |\mathcal{H}|/2^{2m}$. □

Lemma 3. *Let $\mathcal{H}$ be the hash family defined above. Then for any nonzero element $s \in S$ and for any $b \in GF(2)^m$, $|\{h \in \mathcal{H} \ : \ h(s) = b\}| = |\mathcal{H}|/2^m$.*

Proof. clear. □

2.2 Interactive Hashing

Interactive hashing is a two-party protocol between a challenger Alice and a responder Bob. Cachin, Crépeau, and Marcil [4] gave a new elegant analysis on it in order to be used to construct OT in the bounded-storage model. Bob has a secret t-bit string $\chi \in T \subset GF(2)^t$, where $|T| \leq 2^{t-k}$ and χ and T are unknown to Alice. At the end of the protocol, Alice receives two strings, one of which is χ, but Alice does not know which one is χ. Also, Bob cannot force both two strings to be in T, except with a small probability $\nu(k)$.

The following interactive hashing protocol is proposed in Noar, Ostrovsky, Venkatesan and Yung [15]: Alice randomly chooses $t-1$ linearly independent vectors $a_1, \cdots, a_{t-1} \in GF(2)^t$. The protocol then proceeds in $t-1$ rounds. In Round i, for each $i = 1, \cdots, t-1$,

1. Alice announces a_i to Bob.
2. Bob computes $b_i = a_i \cdot \chi$ and sends b_i to Alice.

At the end, both Alice and Bob have the same system of linear equations $b_i = a_i \cdot \chi$, $i = 1, \cdots, t-1$ over $GF(2)$. Since $a_1, \cdots, a_{t-1} \in GF(2)^t$ are linearly independent, the system has exactly two t-bit strings χ_1, χ_2 as solutions and one of them is χ by standard linear algebra. Thus Alice does not know information-theoretically that which solution is χ. Also, the condition that Bob cannot force both two strings to be in T, except with a small probability $\nu(k)$, was proved in [4].

Since the round and communication complexities of NOVY protocol, which transmits the string of t bits, are $t-1$ rounds and t^2-1 bits respectively, the protocol which is based on NOVY protocol has very large round and communication complexities.

2.3 Extended Interactive Hashing Protocol

We propose a new scheme between a challenger Alice with no input and a responder Bob with input string χ which provides a way to isolate more than two strings. Bob has a secret t-bit string $\chi \in T \subset GF(2)^t$, where $|T| \leq 2^{t-k}$ and χ and T are unknown to Alice. For some positive integers l and m, let $t = lm$. The protocol should meet the following requirements:

1. Bob sends a secret t-bit string in such a way that Alice receives 2^m t-bit strings and one of them is χ, but Alice does not know that which one is χ.
2. Bob cannot force any two of them to be in T, except with a small probability $\nu(k)$.

We regard $GF(2)^t$ as $(GF(2)^m)^l$ and let $S = (GF(2)^m)^l$. Bob chooses a secret t-bit string $\chi = (\chi_{l-1}, \cdots, \chi_1, \chi_0) \in S$, where $\chi_i \in GF(2)^m, 0 \leq i \leq l-1$. Now we consider the universal family $\mathcal{H}$ of hash functions from S to $GF(2)^m$ which is defined in Section 2.1 as

$$\mathcal{H} = \{h_\zeta \ : \ \zeta = (\zeta_{l-1}, \cdots, \zeta_1, \zeta_0) \in S\},$$

where h_ζ is defined in (1).

Our scheme is described below.

Protocol : The protocol operates in $t/m-1$ rounds. In Round i, for $i = 1, \cdots, t/m-1$,

1. Alice chooses a function $h_i \in \mathcal{H}$ with uniform distribution. Let $a_i \in GF(2)^t$ be the description vector of h_i such that $h_i = h_{a_i}$. If a_i is linearly dependent in $a_1, \cdots, a_{i-1}$, then Alice repeats this step until it is independent. Alice sends a_i to Bob.
2. Let $a_i = (a_i^{(t/m-1)}, \cdots, a_i^{(1)}, a_i^{(0)}) \in S$, $\; a_i^{(j)} \in GF(2)^m$, $\; 0 \le j \le t/m-1$. Bob computes m-bit $b_i = h_{a_i}(\chi) = \phi^{-1}\left(\sum_{j=0}^{t/m-1} \phi(a_i^{(j)}) \cdot \phi(\chi_j)\right)$, and sends b_i to Alice.

After the $t/m-1$ rounds, both Alice and Bob have the same $t/m-1$ linear equations over $GF(2)^m$ with χ as a solution. The system has exactly 2^m t-bit strings $\chi_0, \cdots, \chi_{2^m-1}$ as solutions, one of which is χ. We call this scheme *extended interactive hashing.* We note that in case of $m = 1$, our protocol is the same as interactive hashing.

It is clear that Alice does not know information-theoretically that which solution is χ. Thus Condition 1 of extended interactive hashing is satisfied. We now come to Condition 2 regarding the security against a dishonest responder Bob. In our protocol, Bob can cheat if he can answer Alice's queries in such a way that T contains two distinct elements s_1, s_2 received by Alice. In Theorem 1, we show that Bob can only cheat in extended interactive hashing if the size of $|T|$ is close to $|GF(2)^t| = 2^t$. In order to prove this, we need some lemmas.

The following lemma shows that each round of scheme reduces the size of T by a factor of almost 2^m with very high probability. This approach was used first to prove the security of interactive hashing in [4]. We improve this method in our model.

Lemma 4. *Let $T \subset GF(2)^t$ be any subset with $|T| = 2^{\alpha t}$ for $0 < \alpha < 1$ and let p be a positive integer such that $p \le \alpha t/3$. Let m be a positive integer which is a divisor of t. Let $\mathcal{H}$ be the universal family of hash functions from $GF(2)^t$ to $GF(2)^m$ defined above. Let U be a random variable with uniform distribution over $\mathcal{H}$. Then for any $b \in GF(2)^m$,*

$$\Pr\left[\; |\{s \in T :\; U(s) = b\}| < \left(\frac{1}{2^m} + \frac{1}{2^{p+m/2}} + \frac{1}{2^{3p}}\right)|T| \;\right] \;\ge\; 1 - 2^{-p}.$$

Proof. For any $s \in T$ and $b \in GF(2)^m$, we consider the following random variables

$$X_{(b,s)} = \begin{cases} 1 & \text{if } U(s) = b \\ 0 & \text{otherwise} \end{cases}$$

and their sum $X_b = \sum_{s \in T} X_{(b,s)} = |\{s \in T :\; U(s) = b\}|$. Thus we must show that for any $b \in GF(2)^m$,

$$\Pr\left[X_b \;<\; \left(\frac{1}{2^m} + \frac{1}{2^{p+m/2}} + \frac{1}{2^{3p}}\right)|T|\;\right] \;\ge\; 1 - 2^{-p}. \tag{3}$$

Case 1: $b \neq 0 \in GF(2)^m$.
By the definition $X_{(b,s)}$ and Lemma 3, we obtain that for any $s \neq 0 \in T$

$$E[X_{(b,s)}] = E[X^2_{(b,s)}] = 1 \cdot \frac{|\{h \in \mathcal{H} :\ h(s) = b\}|}{|\mathcal{H}|} = \frac{1}{2^m},$$

and $X_{(b,0)} = X^2_{(b,0)} = 0$ by the definition of our hash family. Thus $E[X_b] = \frac{|T|-1}{2^m}$. By the definition of X_b, we obtain that

$$E[X_b^2] = E[\sum_{s \in T} X^2_{(b,s)} + 2 \sum_{s_i < s_j \in T} X_{(b,s_i)} X_{(b,s_j)}].$$

Since $b \neq 0$, $X_{(b,0)} = 0$. Using this fact and Lemma 2 we obtain that

$$\begin{aligned} E[\sum_{s_i < s_j \in T} X_{(b,s_i)} X_{(b,s_j)}] &= \sum_{0 < s_i < s_j \in T} E[X_{(b,s_i)} X_{(b,s_j)}] \\ &= \sum_{0 < s_i < s_j \in T} \frac{\{h \in \mathcal{H} :\ h(s_i) = h(s_j) = b\}}{|\mathcal{H}|} \\ &< \frac{(|T|-1)^2}{2} \cdot (\frac{1}{2^m})^2. \end{aligned}$$

Thus, we have $E[X_b^2] < \frac{|T|-1}{2^m} + \left(\frac{|T|-1}{2^m}\right)^2$ and

$$\begin{aligned} Var[X_b] &= E[X_b^2] - (E[X_b])^2 \\ &< \frac{|T|-1}{2^m}. \end{aligned}$$

Now, by Chebychev Inequality we obtain that for any $b \neq 0 \in GF(2)^m$ and $\delta > 0$

$$\Pr\left[\left|X_b - \frac{|T|-1}{2^m}\right| \geq \delta\right] < \frac{|T|-1}{2^m \delta^2}.$$

Substituting $\delta = \sqrt{2^p(|T|-1)/2^m}$, we have

$$\Pr\left[\left|X_b - \frac{|T|-1}{2^m}\right| \geq 2^{\frac{p+\alpha t - m}{2}}\right] < 2^{-p}.$$

Hence, if $p \leq \alpha t/3$, then with probability at least $1 - 2^{-p}$, we obtain

$$\begin{aligned} X_b &< \left(\frac{1}{2^m} + 2^{\frac{p+\alpha t - m}{2} - \alpha t}\right) |T| \\ &< \left(\frac{1}{2^m} + \frac{1}{2^{p+m/2}}\right) |T| \end{aligned}$$

and (3) is satisfied.

Case 2: $b = 0 \in GF(2)^m$.
Using $X_{(0,0)} = 1$, Lemma 2 and Lemma 3, we obtain that $E[X_0] = \frac{|T|-1}{2^m} + 1$ and $E[X_0^2] < \frac{3(|T|-1)}{2^m} + 1 + \left(\frac{|T|-1}{2^m}\right)^2$. Thus $Var[X_0] < \frac{|T|-1}{2^m}$. By Chebychev Inequality we obtain that for any $\delta > 0$

$$\Pr\left[\left|X_0 - \left(\frac{|T|-1}{2^m} + 1\right)\right| \geq \delta\right] < \frac{|T|-1}{2^m \delta^2}.$$

Substituting $\delta = \sqrt{2^p(|T|-1)/2^m}$, we have that with probability at least $1-2^{-p}$,

$$X_0 < \left(\frac{1}{2^m} + \frac{1}{2^{p+m/2}} + \frac{1}{2^{3p}}\right)|T|$$

using $p \leq \alpha t/3$, and the lemma is proved. □

The following lemma was proved in [4]

Lemma 5. [4] *Let $T \subset GF(2)^t$ be any subset with $|T| = 2^{\alpha t}$ for $0 < \alpha < 1$. Let p and q be positive integers such that $2\alpha t < mq - p$ and $p, mq \leq t$ where m is a divisor of t. Let $\mathcal{H}$ be the universal family of hash functions from $GF(2)^t$ to $GF(2)^{mq}$. Let U be a random variable with uniform distribution over $\mathcal{H}$. Then for any distinct $s_1, s_2 \in T$, we have*

$$\Pr[U(s_1) = U(s_2)] \leq 2^{-p}.$$

Lemma 6. *Suppose that Alice and Bob engage in extended interactive hashing of a t-bit string as described above. Let $T \subset GF(2)^t$ be any subset with $|T| = 2^{\alpha t}$ for $0 < \alpha < 1$ and let r be a positive integer such that $\log_2 t \leq r \leq \alpha t/6$. Let m be a positive integer which is a divisor of t and $m \leq 2r$. If $\alpha < 1 - \frac{8r+2m+2}{t}$, then with probability at most $\frac{1}{m2^r}$, Bob can answer Alice's queries in such a way that Bob's answers are consistent for two distinct elements $s_1, s_2 \in T$.*

Proof. For $i = 1, \cdots, t/m-1$, let $T_i \subset T$ be the subset of T satisfying $h_j(s) = b_j$, for $j = 1, \cdots, i$, after Round i of the extended interactive hashing protocol. Let $p = 2r$. Then using $r \leq \alpha t/6$ and $\alpha < 1 - \frac{8r+2m+2}{t}$, we obtain that $\alpha t \geq 3p$ and $\frac{\alpha t - 3p}{m} + 1 < \frac{t}{m} - 1$. Thus there exists a positive integer i_j such that

$$0 \leq \frac{\alpha t - 3p}{m} < i_j \leq \frac{\alpha t - 3p}{m} + 1 < \frac{t}{m} - 1. \tag{4}$$

Applying Lemma 4 by induction on i from 1 to $i_j - 1$, we get

$$|T_i| < \left(\frac{1}{2^m} + \frac{1}{2^{p+m/2}} + \frac{1}{2^{3p}}\right)^i |T|,$$

except with probability at most $i \cdot 2^{-p}$. Thus, we obtain that

$$\log_2 |T_{i_j}| < (\alpha t - m i_j) + i_j \log_2(1 + 2^{m/2-p} + 2^{-3p+m}) < 3p + 1 \tag{5}$$

by (4) and $i_j \log_2(1 + 2^{m/2-p} + 2^{-3p+m}) < t/m \cdot (2^{m/2-p} + 2^{-3p+m}) < 1$.

Now we want to apply Lemma 5 for step i_j (round i_j through $t/m - 1$ collectively) using T_{i_j}. Since $\alpha < 1 - \frac{4p+2m+2}{t}$, $4p < t - \alpha t - 2m - 2$ and $2\log_2 |T_{i_j}| < 6p + 2 < 2p + t - \alpha t - 2m$ by (5). Using (4) we get

$$2p + t - \alpha t - 2m = t - m\left(\frac{\alpha t - 3p}{m} + 2\right) - p \le t + m(-i_j - 1) - p$$

and $2\log_2 |T_{i_j}| < m(t/m - 1 - i_j) - p$ holds. Hence we can apply Lemma 5 and the overall failure probability is at most $(i_j + 1)2^{-p} < t/m \cdot 2^{-p} < \frac{1}{m2^r}$, which proves the lemma. □

The following theorem shows that Condition 2 of extended interactive hashing is satisfied.

Theorem 1 *Suppose that Alice and Bob engage in extended interactive hashing of a t-bit string as defined above. For positive integers l and m, let $t = lm$. Let $T \subset GF(2)^t$ be any subset with $|T| \le 2^{t-k}$ where k satisfies $\log_2 t \le k \le 2t/3$. If $m < \frac{k-2}{6}$, then with probability at most $\frac{2^{-O(k)}}{m}$, Bob can answer Alice's queries in such a way that Bob's answers are consistent for two distinct elements $s_1, s_2 \in T$.*

Proof. For any positive integer r which satisfies $\log_2 t \le r \le (t-2)/18$, let $k = 12r + 2$. Then we get $r \le \frac{t-k}{6}$ and $m < 2r$. Thus the theorem follows from Lemma 6. □

Corollary 1 *Suppose that Alice and Bob engage in extended interactive hashing of a t-bit string as defined above. For positive integers l and m, let $t = lm, m < t$. Let $T_0, T_1 \subset GF(2)^t$ be any two subsets with $|T_0|, |T_1| \le 2^{t-k}$ where k satisfies $\log_2 t \le k \le 2t/3$. If $m < \frac{k-2}{6}$, then the probability that Bob can answer Alice's queries such that two distinct elements, which one lies in T_0 and the other one lies in T_1, are consistent with his answers is at most $\frac{2^{-O(k)}}{m}$.*

3 1-out of-N Oblivious Transfer Protocol

In this section we describe an efficient protocol for OT_1^N in the bounded-storage model. Throughout the paper, let k be a security parameter and M be the length of a public random string, and let $L = \tau M$, $\tau < 1$, be the storage bound on the reciever Bob. For simplicity, we only consider $L = M/6$(i.e. $\tau = 1/6$). For any $\tau < 1$ we can obtain similar results.

An OT_1^N scheme is a two-party protocol between the sender Alice who possesses N secret bits $X_0, \cdots, X_{N-1} \in GF(2)$ and the reciver Bob who would like to learn one of them at his choice. We assume that Alice is honest, that is, it won't send secrets that are not claimed. An OT_1^N scheme should satisfy the following requirements :

1. Correctness : if Alice and Bob follow the protocol, Bob obtains X_c after executing the protocol, where $c \in \{0, \cdots, N-1\}$ is a secret value of his choice.

2. Bob's privacy : after executing the protocol with Bob, Alice shall not get any information about Bob's secret value c.
3. Alice's privacy : after executing the protocol with Alice, Bob does not learn any information about other secrets $X_i, i \neq c$ or their combination except with a negligible probability $\nu(k)$.

3.1 Basis Ideas

In this subsection we explain the basic ideas of our protocol for OT_1^N. Let $n = 2\sqrt{kM}$.

First, Alice and Bob choose independent random subsets $\mathcal{A}$, $\mathcal{B} \subset \{1, \cdots, M\}$ with $|\mathcal{A}| = |\mathcal{B}| = n$, respectively. If public random string $\alpha \xleftarrow{R} GF(2)^M$ is broadcasted, Alice stores $\alpha[i], \forall i \in \mathcal{A}$ and Bob stores $\alpha[j], \forall j \in \mathcal{B}$, where $\alpha[i]$ is the i-th bit of α. Then Alice sends her subset $\mathcal{A}$ to Bob, and Bob computes $\mathcal{A} \cap \mathcal{B}$. Following lemma shows that $|A \cap B| \geq k$ with very high probability.

Lemma 7. [10] Let $\mathcal{A}, \mathcal{B}$ be two independent random subset of $\{1, \cdots, M\}$ with $|\mathcal{A}| = |\mathcal{B}| = 2\sqrt{kM}$. Then $\Pr[\ |\mathcal{A} \cap \mathcal{B}| < k\ ] < e^{-k/4}$.

Fact 1 (Encoding k-Element Subsets) [4] Each of the $\binom{n}{k}$ k-element subsets of $\{1, \cdots, n\}$ can be uniquely encoded as a $\lceil \log_2 \binom{n}{k} \rceil$-bit string.

Next, Bob encodes a random k-element subset $\mathcal{A}_I \subset \mathcal{A} \cap \mathcal{B}$ as a $\lceil \log_2 \binom{n}{k} \rceil$-bit string and sends $\mathcal{A}_I$ to Alice by the extended interactive hashing protocol defined in Section 2.3. After executing the extended interactive hashing protocol between Alice and Bob, they can construct one "good" set and $N-1$ "bad" sets. Bob knows the "good" set, but does not learn any information about the "bad" sets. Alice knows all of the sets, but does not distinguish between the "good" set and the"bad" sets.

Next, Bob asks Alice to encrypt X_c with the "good" set and other secrets $X_i, i \neq c$ with the bad sets. Since Bob knows the "good" set, not the "bad" sets, he can recover X_c, but not $X_i \neq c$.

3.2 Protocol for OT_1^N

We propose the OT_1^N protocol for a receiver with bounded memory size. The protocol uses N public random string $\alpha_0, \cdots, \alpha_{N-1} \xleftarrow{R} GF(2)^M$. Let $n = 2\sqrt{kM}$ and let $t = \lceil \log_2 \binom{n}{k} \rceil$. For some positive integers l and m, suppose $t = lm$ and $m < (k-2)/6$.

Protocol (OT_1^N) : A sender Alice has N input bits $X_0, \cdots, X_{N-1}$ when $N = 2^u, 1 \leq u \leq m$. A receiver Bob chooses $c \in \{0, \cdots, N-1\}$ and want to know X_c.

1. Alice randomly chooses N sets $\mathcal{A}^{(0)} = \{a_1^{(0)}, \cdots, a_n^{(0)}\}, \cdots, \mathcal{A}^{(N-1)} = \{a_1^{(N-1)}, \cdots, a_n^{(N-1)}\} \subset \{1, \cdots, M\}$ with length n. Bob randomly chooses N sets

$\mathcal{B}^{(0)} = \{b_1^{(0)}, \cdots, b_n^{(0)}\}, \cdots, \mathcal{B}^{(N-1)} = \{b_1^{(N-1)}, \cdots, b_n^{(N-1)}\} \subset \{1, \cdots, M\}$ with length n.

2. If the first public random string $\alpha_0 \xleftarrow{R} GF(2)^M$ is broadcasted, Alice stores $\alpha_0[a_1^{(0)}], \cdots, \alpha_0[a_n^{(0)}]$ and Bob stores $\alpha_0[b_1^{(0)}], \cdots, \alpha_0[b_n^{(0)}]$. After a short time, if the second public random string $\alpha_1 \xleftarrow{R} GF(2)^M$ is broadcasted, Alice stores $\alpha_1[a_1^{(1)}], \cdots, \alpha_1[a_n^{(1)}]$ and Bob stores $\alpha_1[b_1^{(1)}], \cdots, \alpha_1[b_n^{(1)}]$. After iterative procedures, if $\alpha_{N-1} \xleftarrow{R} GF(2)^M$ is broadcasted, Alice stores the $\alpha_{N-1}[a_1^{(N-1)}], \cdots, \alpha_{N-1}[a_n^{(N-1)}]$ and Bob also stores the $\alpha_{N-1}[b_1^{(N-1)}], \cdots, \alpha_{N-1}[b_n^{(N-1)}]$.
3. Alice sends $\mathcal{A}^{(0)}, \cdots, \mathcal{A}^{(N-1)}$ to Bob. Bob randomly chooses $\varepsilon \xleftarrow{R} \{0, \cdots, N-1\}$, and computes $\mathcal{A}^{(\varepsilon)} \cap \mathcal{B}^{(\varepsilon)}$. If $\mid \mathcal{A}^{(\varepsilon)} \cap \mathcal{B}^{(\varepsilon)} \mid < k$, then he aborts the protocol. Otherwise, Bob chooses a set $I = \{i_1, \cdots, i_k\}$ such that $\mathcal{A}_I^{(\varepsilon)} = \{a_{i_1}^{(\varepsilon)}, \cdots, a_{i_k}^{(\varepsilon)}\} \subset \mathcal{A}^{(\varepsilon)} \cap \mathcal{B}^{(\varepsilon)}$.
4. Bob encodes I as a t-bit string, where $t = \lceil \log_2 \binom{n}{k} \rceil$. Bob sends I to Alice with the extended interactive hashing protocol in $t/m - 1$ rounds. After executing the extended interactive hashing, both Alice and Bob have exactly 2^m t-bit strings, one of which is I. Bob chooses N subsets $I_0 < \cdots < I_{N-1}$ such that $I = I_\delta$ for some $\delta \in \{0, \cdots, N-1\}$ and such that N strings that encode $I_0, \cdots, I_{N-1}$ are among the 2^m possible strings from the extended interactive hashing protocol, and sends them to Alice.
5. Alice checks whether N k-subsets $I_0 < \cdots < I_{N-1} \subset \{1, \cdots, n\}$ received in Step 4 are contained in all of 2^m k-subsets, computed by the extended interactive hashing protocol. If any one of N k-subsets isn't contained in 2^m k-subsets, she aborts the protocol. For some $\delta \in \{0, \cdots, N-1\}$, $I = I_\delta$. Bob knows δ, but Alice does not know δ .
6. Bob sends u bits $\gamma = \delta \oplus \varepsilon$ and $\rho = c \oplus \varepsilon$ to Alice, where for any $x, y \in \{0, \cdots, N-1\}$, $x \oplus y$ is defined as follows: $x \oplus y = (x_0 \oplus y_0, \cdots, x_{u-1} \oplus y_{u-1}) \in GF(2)^u$ where $x = (x_0, \cdots, x_{u-1}), y = (y_0, \cdots, y_{u-1}) \in GF(2)^u$.
7. Alice sets $Y_0 = \bigoplus_{j=1}^k \alpha_0[a_{I_\gamma[j]}^{(0)}], \cdots, Y_{N-1} = \bigoplus_{j=1}^k \alpha_{N-1}[a_{I_{\gamma \oplus N-1}[j]}^{(N-1)}]$ where $I_l[j]$ denote the j-th element of k-subset I_l, for $l = 0, \cdots, N-1$. Then Alice computes $Z_0 = X_0 \bigoplus Y_\rho, \cdots, Z_{N-1} = X_{N-1} \bigoplus Y_{\rho \bigoplus N-1}$, and sends $Z_0, \cdots, Z_{N-1}$ to Bob.
8. Bob gets $X_c = Z_c \bigoplus Y_\varepsilon$.

Remark 1. Alice and Bob store $N \cdot n = 2N\sqrt{kM}$ bits in Step 2. Alice and Bob also store t^2/m bits in the extended interactive hashing of the Step 4. Here $t = \lceil \log_2 \binom{n}{k} \rceil < k \cdot (\log_2 n - \log_2 k/e)$. Because $k \ll M$, they need to store $O(n)/m$ bits. Thus, in order to implement the protocol, Alice and Bob should store $N \cdot n + O(n)/m$ bits.
Remark 2. The probability that an honest receiver Bob aborts in Step 3 of the protocol, is not more than $e^{-k/4}$ by Lemma 6.

Correctness : Since $Y_\varepsilon = \bigoplus_{j=1}^{k} \alpha_\varepsilon[a^{(\varepsilon)}_{I_{\gamma\oplus\varepsilon}[j]}] = \bigoplus_{j=1}^{k} \alpha_\varepsilon[a^{(\varepsilon)}_{I[j]}]$, Bob can know Y_ε. Thus, he can compute $X_c = Z_c \bigoplus Y_{\rho\oplus c} = Z_c \bigoplus Y_\varepsilon$.

Bob's Privacy : Because Alice does not know ε defined in Step 3 and δ defined in Step 5, She gains no information about the Bob's secret c with γ and ρ received from Step 6.

Alice's privacy : In order to prove the security against a dishonest receiver Bob, who can store $L = M/6$ bits, we apply the method of proof in the Ding's model [10]. If α_0 is broadcasted in Step 2, Bob computes an arbitrary function $\eta_0 = A_0(\eta_0)$, $|\eta_0| = M/6$ using unlimited computing power. And if α_1 is broadcasted, Bob computes an arbitrary function $\eta_1 = A_1(\eta_0, \alpha_1)$, $|\eta_1| = M/6$. After iterative procedures, if α_{N-1} is broadcasted, Bob computes an arbitrary function $\eta_{N-1} = A_{N-1}(\eta_{N-2}, \alpha_{N-1})$, $|\eta_{N-1}| = M/6$. In Step 3 - Step 6, using $\mathcal{A}^{(0)}, \cdots, \mathcal{A}^{(N-1)}$ and η_{N-1}, Bob uses an arbitrary strategy in interacting with Alice. After executing the protocol, Bob tries to gain an information about $X_i, i \neq c$, using the information η_{N-1} on $(\alpha_0, \cdots, \alpha_{n-1})$, $Z_0, \cdots, Z_{N-1}$ received from Alice in Step 7, and all information Ω which he gains in Step 3 - Step 6.

Theorem 2 *Consider the OT_1^N protocol defined above. For any $A_0 : GF(2)^M \longrightarrow GF(2)^{M/6}$, $A_1 : GF(2)^{M/6} \times GF(2)^M \longrightarrow GF(2)^{M/6}$, $\cdots$, $A_{N-1} : GF(2)^{M/6} \times GF(2)^M \longrightarrow GF(2)^{M/6}$, for any strategy Bob uses in Step 3 - Step 6 of the protocol, with probability at least $1 - 2^{-O(k)} - N \cdot 2^{-0.02M}$, there exist some $\rho \in \{0, 1, \cdots, N-1\}$ such that $\forall X_0, \cdots, X_{N-1} \in GF(2)$, $\forall c \in \{0, \cdots, N-1\}$, $\forall i \in \{1, \cdots, N-1\}$ and for any distinguisher $\mathcal{D}$,*

$$\begin{aligned} &|\ Pr[\ \mathcal{D}(\eta_{N-1}, \Omega, Y_{\rho\oplus i} \oplus X_c, Y_\rho \oplus X_{c\oplus i}) = 1\ ] \\ &\quad -\ Pr[\ \mathcal{D}(\eta_{N-1}, \Omega, Y_{\rho\oplus i} \oplus X_c, Y_\rho \oplus 1 \oplus X_{c\oplus i}) = 1\ ]\ | < 2^{-k/3}, \end{aligned} \tag{6}$$

where $\eta_0 = A_0(\alpha_0)$, $\eta_1 = A_1(\eta_0, \alpha_1)$, $\cdots$, $\eta_{N-1} = A_{N-1}(\eta_{N-2}, \alpha_{N-1})$, Ω denotes all the information Bob obtains in Step 3 - Step 6, and $Y_0, \cdots, Y_{N-1}$ are defined in Step 7. Thus the view of Bob is essentially the same, even though $X_{c\oplus i}$ is replaced by $1 \oplus X_{c\oplus i}$. Hence Bob gains no information about any non-trivial function or relation involving more than two X_i's in the protocol.

A proof of this theorem which guarantees the privacy of Alice is given in the appendix.

4 Complexity

In the bounded-storage model, complex of OT_1^N mainly depends on the extended interactive hashing scheme. Since the complexity of the extended interactive hashing scheme for OT_1^N is similar to that of OT_1^2, we compare the complexity of our extended interactive hashing protocol for OT_1^2 with the complexity of NOVY protocol, which is an interactive hashing scheme used in the CCM protocol [4] and Ding's protocol [10].

The NOVY protocol, which transmits the string of t-bits, has $t-1$ rounds complexity and $(t-1)\cdot(t+1) = t^2-1$ bits of communication complexity. On the other hand our extended interactive hashing protocol has $t/m-1$ rounds complexity and $(t/m-1)\cdot(t+m) = t^2/m-m$ bits of communication complexity when m divides t. In case of $m=1$, we note that our protocol and the NOVY protocol are same. If there exists m such that $m>1$, our protocol can be constructed about m times as efficient as compared with the NOVY protocol. As m is large , we see that the complexity of our protocol is more reduced. By Theorem 1, m satisfies the following condition ; $1 \le m < (k-2)/6$, where k is a security parameter. Thus if we choose the largest integer m such that m divides t and $1 \le m < (k-2)/6$, then we can obtain the integer m which makes our protocol most efficient. For example, assume that the length of public random string is Petabit (i.e. $M = 10^{15}$) and $1000 \le k \le 10000$ for a security parameter k. Table 1 gives the information for a security parameter k that we can choose in our protocol.

Table 1. $M = 10^{15}, n = \lceil 2\sqrt{kM} \rceil, t = \lceil \log_2 \binom{n}{k} \rceil$ and m_{max} is the largest positive integer m, which divides t and $m < (k-2)/6$.

k	the number of k such that $m_{max} \ge \sqrt{t}$	the number of k such that $m_{max} = 1$
1000 - 2000	218	101
2001 - 3000	329	100
3001 - 4000	353	92
4001 - 5000	389	95
5001 - 6000	403	90
6001 - 7000	414	77
7001 - 8000	440	75
8001 - 9000	426	93
9000 - 10000	445	65

In case $m_{max} = 1$ in Table 1, our interactive hashing protocol is simply equivalent to the NOVY protocol. By Table 1 we have that the number of k such that $m_{max} = 1$ is less than 10% for $1000 \le k \le 10000$. If we choose k such that $m_{max} \ge \sqrt{t}$, then we can construct protocol which has much lower communication complexity of $t^{3/2} - t^{1/2}$ bits than that of the NOVY protocol. Such k are more than 20% for $1000 \le k \le 2000$, 30% for $2001 \le k \le 5000$ and 40% for $5001 \le k \le 10000$. Hence, we can choose k easily such that our extended interactive hashing for OT_1^2 becomes more efficient than the NOVY protocol for CCM protocol and Ding's protocol.

5 Conclusion

In this paper we propose the OT_1^N protocol as a generalization of the Ding's protocol for OT_1^2 in the bounded-storage model. Furthermore, when $N = 2$, our protocol is similar to that of Ding, but is constructed more efficient than that of Ding. We used the efficiently extended interactive hashing protocol for the sake of reducing a complexity of the protocol. The proposed extended interactive hashing protocol which transmits t-bit string has $t/m - 1$ round complexity and $(t/m - 1) \cdot (t + m) = t^2/m - m$ bits of communication complexity when m divides t, and provides a way to isolate more than two strings. We note that a given m in this paper must divide t and satisfy $m < (k - 2)/6$. And we show that we can choose an integer m such that the protocol has $t^{3/2} - t^{1/2}$ bit communication compexity which is much lower than that of NOVY protocol by a concrete example. This fact provides a partial answer for an open problem raised in [10]. Using such extended interactive hashing, we also constructed the protocol for OT_1^N having almost the same efficiency as OT_1^2 scheme.

References

1. C. H. Bennett, G. Brassard, C. Crépeau, and M. H. Skubiszewska, *Practical quantum oblivious transter protocols*, In Advances in Cryptology - CRYPTO '91, pp. 351-366, 1991.
2. G. Brassard, C. Crépeau, and J. M. Robert, *All-or-nothing disclosure of secrets*, In Advances in Cryptology - Crypto 86, pp. 234-238, 1987.
3. C. Cachin and U. Maurer, *Unconditional security against memory-bounded adversaries*, In Advances in Cryptology - CRYPTO '97, pp. 292-306, 1997.
4. C. Cachin, C, Crépeau, and J. Marcil, *Oblivious tranfser with a memory-bounded receiver*, In Proc. 39th IEEE Symposium in Foundations of Computer Science, pp. 493-502, 1998.
5. J. L. Carter and M. N. Wegman, *Universal classes of hash functions*, Journal of Computer and System Sciences 18, pp. 143-154, 1979.
6. C. Crépeau, *Equivalence between two flavours of oblivious transter*, In Advances in Cryptology - CRYPTO '87, pp. 351-368, 1987.
7. C. Crépeau, J. van de Graff, and A. Tapp, *Committed oblivious transter and private multy-party computations*, In Advances in Cryptology - CRYPTO '95, pp. 110-123, 1995.
8. C. Crépeau and J. Kilian, *Achieving oblivious transfer using weakened security assumptions*, In Proc. 29th IEEE Symposium in the Foundatins of Computer Science, pp. 42-52, 1988.
9. C. Crépeau and M. Sántha, *On the reversibility of oblivious transter and private multi-party computations*, In Advances in Cryptology - CRYPTO '95, pp. 110-123, 1995.
10. Y. Z. Ding, *Oblivious Transfer in the Bounded Storage Model*, In Advances in Cryptology - CRYPTO 2001, pp. 155-170, 2001.
11. S. Even, O. Goldreich, and A. Lempel, *A randomized protocol for signing contracts*, In Advances in Cryptology - CRYPTO '82, pp. 205-210, 1982.
12. O. Goldreich, S. Micali, and A. Wigderson, *How to play ahy mental game or a completeness theorem for protocols with homest majority*, In Proc. 19th ACM Symposium on Theory of Computing, pp. 218-229, 1987.

13. J. Kilian,*Founding cryptography on oblivious transfer*, In Proc. 20th ACM Symposium on Theory of Computing, pp. 20-31, 1988.
14. J. Kilian, E. Kushilevitz, S. Micali, and R. Ostrovsky, *Reducibility and completeness in private computations*, SIAM Journal on Computing, 29(4), pp.1189-1208, 2000.
15. M. Naor, R. Ostrovsky, R. Venkatesan, and M. Yung, *Perfect sero-knowledge arguments for NP using any one-way function*, Journal of Cryptology, 11(2), pp. 87-108, 1998. Preliminary version presented at CRYPTO '92.
16. M. O. Rabin, *How to exchange secrets by oblivious transfer*, Technical Report TR-81, Harvard University, 1981.
17. M. O. Rabin, *Transaction Protection by Beacons*, JCSS 27(2), pp. 256-267, 1983.
18. P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Computing, 26(5), pp. 1484-1509, 1997.
19. D. R. Stinson, *On the connections between universal hasing, combinatorial designs and error-correcting codes*, Congressus Numerantium 114, pp. 7-27, 1996.
20. D. R. Stinson *Universal hash families and the leftover hash lemma, and applications to crypto graphy and computing*, 2002. preprint, see, http://cacr.math.uwaterloo.ca/ dstinson/
21. A. C. Yao, *How to generate and exchange secrets*, In Proc. 27th IEEE Symposium on the Foundations of Computer Science, pp. 162-167, 1986.

A Proof of Theorem 2

We extend the proof in Ding [10] to deal with OT_1^N. We use the same definitions and lemmas as given in [10].

Definition 2 *Define* $\mathcal{K} \stackrel{\text{def}}{=} \{I \subset \{1, \cdots, M\} : |I| = k\}$.

Definition 3 *Let* $E \subset GF(2)^M$ *and* $I \in \mathcal{K}$. *We say that* I *is good for* E *if*

$$\left| \frac{|\{\alpha \in E : \bigoplus_{i=1}^{k} \alpha[I[i]] = 0\}|}{|E|} - \frac{|\{\alpha \in E : \bigoplus_{i=1}^{k} \alpha[I[i]] = 1\}|}{|E|} \right| < 2^{-k/3}.$$

Definition 4 *Let* $E \subset GF(2)^M$. *We say that* E *is fat if* $|E| \geq 2^{0.813M}$.

Lemma 8. [10] *For any function* $f : GF(2)^M \to GF(2)^{M/6}$ *and* $\alpha \xleftarrow{R} GF(2)^M$,

$$\Pr[f^{-1}(f(\alpha)) \text{ is fat }] > 1 - 2^{-0.02M}.$$

Definition 5 *For* $\mathcal{A} \subset \{1, \cdots, M\}$, *define* $\mathcal{K}_{\mathcal{A}} \stackrel{\text{def}}{=} \{I \subset \mathcal{A} : |I| = k\}$.

Definition 6 *For* $\mathcal{A} \subset \{1, \cdots, M\}$ *and* $E \subset GF(2)^M$, *define*

$$\mathcal{B}_E^{\mathcal{A}} \stackrel{\text{def}}{=} \{I \subset \mathcal{K}_{\mathcal{A}} : I \text{ is not good for } E\}.$$

Lemma 9. [10] *Let* $E \subset GF(2)^M$ *be fat. For a uniformly random* $\mathcal{A} \subset \{1, \cdots, M\}$ *with* $|\mathcal{A}| = n$,

$$\Pr\left[|\mathcal{B}_E^{\mathcal{A}}| < |\mathcal{K}_{\mathcal{A}}| \cdot 2^{-k/6} = \binom{n}{k} \cdot 2^{-k/6}\right] > 1 - 2^{-k/6}.$$

PROOF OF THEOREM 2 : In order to show the equation (6) of Theorem 2, it suffices to show that with probability $1 - 2^{-O(k)} - N \cdot 2^{-0.02M}$, there exists $\rho \in \{0, 1, \cdots, N-1\}$ such that for any $i \in \{1, \cdots, N\}$ and for any distinguisher $\mathcal{D}$,

$$|\Pr[\mathcal{D}(\eta_{N-1}, \Omega, Y_{\rho \oplus i}, Y_\rho) = 1] - \Pr[\mathcal{D}(\eta_{N-1}, \Omega, Y_{\rho \oplus i}, Y_\rho \oplus 1) = 1]| < 2^{-k/3}. \quad (7)$$

Here $\eta_0 = A_0(\alpha_0), \eta_1 = A_1(\eta_0, \alpha_1), \cdots, \eta_{N-1} = A_{N-1}(\eta_{N-2}, \alpha_{N-1})$, Ω denotes all the information Bob obtains in Step 3-Step 6, and $Y_0, \cdots, Y_{N-1}$ are defined in Step 7 of the protocol.

Note that as in the proof of Theorem 1 in [10] it suffices to show the equation (7) in the case that Bob's recording functions $A_0, \cdots, A_{N-1}$ are deterministic.

We prove a slightly stronger result that the equation (7) hold even if Bob stores not only η_{N-1}, but also $\eta_0, \eta_1, \cdots, \eta_{N-2}$. Let

$$E_0 \stackrel{\text{def}}{=} \{\alpha \in GF(2)^M : A_0(\alpha) = \eta_0\}, \ E_1 \stackrel{\text{def}}{=} \{\alpha \in GF(2)^M : A_1(\eta_0, \alpha) = \eta_1\},$$
$$\cdots, \ E_{N-1} \stackrel{\text{def}}{=} \{\alpha \in GF(2)^M : A_{N-1}(\eta_{N-2}, \alpha) = \eta_{N-1}\}.$$

After $\eta_0, \cdots, \eta_{N-1}$ are computed in Step 2 of the protocol, Bob can compute $E_0, \cdots, E_{N-1}$ using unlimited computing power. But given $\eta_0, \cdots, \eta_{N-1}$, all Bob knows about $(\alpha_0, \cdots, \alpha_{N-1})$ are that it is uniformly random in $E_0 \times \cdots \times E_{N-1}$. By Lemma 8, for any recording functions $A_0, \cdots, A_{N-1}$ and for $\alpha_0, \cdots, \alpha_{N-1} \stackrel{R}{\longleftarrow} GF(2)^M$,

$$\Pr[\text{All of } E_0, \cdots, E_{N-1} \text{ are fat }] > 1 - N \cdot 2^{-0.02M} \quad (8)$$

Thus, consider the case that all of $E_0, \cdots, E_{N-1}$ are fat.

Let $\mathcal{A}^{(0)}, \cdots, \mathcal{A}^{(N-1)}$ be the random subsets of $\{1, \cdots, M\}$ with $|\mathcal{A}^{(i)}| = n, \forall i \in \{0, 1, \cdots, N-1\}$, which Alice chooses in Step 1 of the protocol. By (8) and Lemma 9, we have that for any $i \in \{1, \cdots, N-1\}$, for $\rho \in \{0, \cdots, N-1\}$, with probability at least $1 - N \cdot 2^{-0.02M} - 2^{-k/6+1}$,

$$|\mathcal{B}_{E_\rho}^{\mathcal{A}^{(\rho)}}|, |\mathcal{B}_{E_{\rho \oplus i}}^{\mathcal{A}^{(\rho \oplus i)}}| < \binom{n}{k} \cdot 2^{-k/6}. \quad (9)$$

Thus consider the case that $\mathcal{B}_{E_\rho}^{\mathcal{A}^{(\rho)}}, \mathcal{B}_{E_{\rho \oplus i}}^{\mathcal{A}^{(\rho \oplus i)}}$ satisfy (9).

For each $\epsilon \in \{0, \cdots, N-1\}$, denote $\mathcal{A}^{(\epsilon)} = \{a_1^{(\epsilon)}, \cdots, a_n^{(\epsilon)}\}$. For $J = \{j_1, \cdots, j_k\} \subset \{1, \cdots, n\}$, denote $\mathcal{A}_J^{(\epsilon)} = \{a_{j_1}^{(\epsilon)}, \cdots, a_{j_k}^{(\epsilon)}\}$. By Definition 5, $\mathcal{A}_J^{(\epsilon)} \in \mathcal{K}_{\mathcal{A}^{(\epsilon)}}$. Define

$$F_\rho \stackrel{\text{def}}{=} \{J \subset \{1, \cdots, n\} : |J| = k \wedge \mathcal{A}_J^{(\rho)} \in \mathcal{B}_{E_\rho}^{\mathcal{A}^{(\rho)}}\};$$
$$F_{\rho \oplus i} \stackrel{\text{def}}{=} \{J \subset \{1, \cdots, n\} : |J| = k \wedge \mathcal{A}_J^{(\rho \oplus i)} \in \mathcal{B}_{E_{\rho \oplus i}}^{\mathcal{A}^{(\rho \oplus i)}}\}.$$

Using (9) and $|F_\rho| = |\mathcal{B}_{E_\rho}^{\mathcal{A}^{(\rho)}}|$, $|F_{\rho \oplus i}| = |\mathcal{B}_{E_{\rho \oplus i}}^{\mathcal{A}^{(\rho \oplus i)}}|$, we have

$$|F_\rho|, |F_{\rho \oplus i}| < \binom{n}{k} \cdot 2^{-k/6}. \quad (10)$$

Consider $I_0, \cdots, I_{N-1}$ defined in Step 5 of the protocol. Let γ be the first u-bit which Bob sends to Alice in Step 6 of the protocol. Then by (8), (9), (10) and Corollary 1 on the extended interactive hashing, we have that for any strategy Bob uses in Step 3 - Step 6, with probability at least $1 - 2^{-O(k)} - N \cdot 2^{-0.02M}$, $I_{\gamma\oplus\rho} \notin F_\rho \vee I_{\gamma\oplus\rho\oplus i} \notin F_{\rho\oplus i}$. WLOG, assume $I_{\gamma\oplus\rho\oplus i} \notin F_{\rho\oplus i}$. Let $Y_\rho = \bigoplus_{j=1}^{k} \alpha_\rho[a^{(\rho)}_{I_{\gamma\oplus\rho}[j]}]$, $Y_{\rho\oplus i} = \bigoplus_{j=1}^{k} \alpha_{\rho\oplus i}[a^{(\rho\oplus i)}_{I_{\gamma\oplus\rho\oplus i}[j]}]$ as defined in Step 7 of the protocol. Since $I_{\gamma\oplus\rho\oplus i} \notin F_{\rho\oplus i}$, by definition $\mathcal{A}^{(\rho\oplus i)}_{I_{\gamma\oplus\rho\oplus i}[j]} \notin \mathcal{B}_{E^{\mathcal{A}^{(\rho\oplus i)}}_{\rho\oplus i}}$. By definition 3 of goodness, for $\alpha_{\rho\oplus i} \xleftarrow{R} E_{\rho\oplus i}$,

$$|\Pr[Y_{\rho\oplus i} = 0] - \Pr[Y_{\rho\oplus i} = 1]| < 2^{-k/3}.$$

Since $(\alpha_\rho, \alpha_{\rho\oplus i}) \xleftarrow{R} E_\rho \times E_{\rho\oplus i}$, Y_ρ and $Y_{\rho\oplus i}$ are independent. Thus for any $b \in GF(2)$,

$$|\Pr[Y_{\rho\oplus i} = 0 \mid Y_\rho = b] - \Pr[Y_{\rho\oplus i} = 1 \mid Y_\rho = b]| < 2^{-k/3}$$

which proves (7) and the proof is done.