Fully Collusion Resistant Traitor Tracing with Short Ciphertexts and Private Keys

Dan Boneh[1*], Amit Sahai[2**], and Brent Waters[3]

[1] Stanford University, dabo@cs.stanford.edu
[2] U.C.L.A., sahai@cs.ucla.edu
[3] SRI International, bwaters@csl.sri.com

Abstract. We construct a fully collusion resistant tracing traitors system with sublinear size ciphertexts and constant size private keys. More precisely, let N be the total number of users. Our system generates ciphertexts of size $O(\sqrt{N})$ and private keys of size $O(1)$. We first introduce a simpler primitive we call *private linear broadcast encryption* (PLBE) and show that any PLBE gives a tracing traitors system with the same parameters. We then show how to build a PLBE system with $O(\sqrt{N})$ size ciphertexts. Our system uses bilinear maps in groups of composite order.

1 Introduction

Traitor tracing systems, introduced by Chor, Fiat, and Naor [10], help content distributors identify pirates. Consider a content distributor who broadcasts encrypted content to N legitimate recipients. Recipient i has secret key K_i that it uses to decrypt the broadcast. As a concrete example, imagine an encrypted satellite radio broadcast that should only be played on certified radio receivers. The broadcast is encrypted using a public broadcasting key BK. Any certified player can decrypt using its embedded secret key K_i. Certified players, of course, could enforce digital rights restrictions such as "do not copy" or "play once".

The risk for the distributor is that a pirate will hack a certified player and extract its secret key. The pirate could then build a pirate decoder that will extract the cleartext content and ignore any relevant digital rights restrictions. Even worse, the pirate could make its pirate decoder widely available so that anyone can extract the cleartext content for themselves. DeCSS, for example, is a widely distributed program for decrypting encrypted DVD content.

This is where traitor tracing systems come in — when the pirate decoder is found, the distributor can run a *tracing* algorithm that interacts with the pirate decoder and outputs the index i of at least one of the keys K_i that the pirate used to create the pirate decoder. The distributor can then try to take legal action against the owner of this K_i.

* Supported by NSF and the Packard Foundation.

** This research was supported by the NSF Cybertrust and ITR Programs, an Alfred P. Sloan Research Fellowship, and a generous equipment grant from Intel.

We give a precise description of traitor tracing systems in Appendix A. For now we give some intuition that will help explain our results. A traitor tracing system consists of four algorithms *Setup*, *Encrypt*, *Decrypt*, and *Trace*. The setup algorithm generates the broadcaster's key BK, a tracing key TK, and N recipient keys $\mathrm{K}_1, \ldots, \mathrm{K}_N$. The encrypt algorithm encrypts the content using BK and the decrypt algorithm decrypts using one of the K_i. The tracing algorithm is the most interesting — it is an algorithm that takes TK as input and interacts with a pirate decoder, treating it as a black-box oracle. It outputs the index $i \in \{1, \ldots, N\}$ of a key K_i that was used to create the pirate decoder.

In this paper we focus on fully collusion resistant traitor tracing systems. That is, systems that remain secure no matter how many keys are at the disposal of the pirate. Existing traitor tracing systems are not designed to handle arbitrary collusions. When the collusion bound t comes close to N, most existing systems require ciphertext size linear in the number of users, which is no better than the trivial traitor tracing system.

Our results. We construct a practical fully collusion resistant traitor tracing system that has sub-linear size ciphertexts. Our system has the following characteristics:

$$\text{ciphertext-length} = O(\sqrt{N}) \qquad \text{and} \qquad \text{private-key-length} = O(1)$$

Furthermore, decryption time is constant (i.e. depends on the security parameter, but not on N). Other properties of this system include: (1) the broadcaster's key BK is public, but the tracer's key TK must be kept secret, (2) the system is black-box traceable, and (3) is designed for stateless pirate decoders [18]. We give a precise definition of these properties in Appendix A. The system uses bilinear groups of composite order introduced in [5].

We prove security of our tracing algorithm using a tracing technique previously used in [4, 23, 18]. To formalize this technique, we introduce a new primitive called *Private Linear Broadcast Encryption*, or PLBE for short, which is conceptually a simpler primitive than traitor tracing. We show that any secure PLBE gives a (black-box) traitor tracing system. Roughly speaking, a PLBE is a broadcast encryption system [13] that can only broadcast to "linear" sets, that is sets of the form $\{i, i+1, \ldots, N\}$ for some $i = 1, \ldots, N+1$. Thus, a PLBE enables the broadcaster to create ciphertexts that can only be decrypted properly under keys $\mathrm{K}_i, \mathrm{K}_{i+1}, \ldots, \mathrm{K}_N$. A broadcast to everyone, for example, is encrypted using $i = 1$. The main security requirement is that the system should be *private* [1]: a ciphertext should reveal no non-trivial information about the recipient set. That is, a broadcast to users $\{i, \ldots, N\}$ should reveal no non-trivial information about i. We give a precise definition in the next section and show that any secure PLBE gives a secure (black-box) traitor tracing system. In the remainder of the paper we focus on constructing a secure PLBE.

Related work. Traitor tracing systems generally fall into two categories: combinatorial, as in [10, 24, 31, 32, 14, 15, 11, 28, 2, 30, 29, 23], and algebraic, as in [21, 4, 25, 20, 12, 22, 34, 9]. The broadcaster's key BK in combinatorial systems can be either secret or public. Algebraic traitor tracing use public-key techniques and are often more efficient than the public-key instantiations of combinatorial schemes. Some systems, including ours, only provide tracing capabilities. Other systems [25, 23, 17, 16, 12] combine tracing with broadcast encryption to obtain trace-and-revoke features — after tracing, the distributor can revoke the pirate's keys without affecting any other legitimate decoder.

Kiayias and Yung [20] describe a black-box tracing system that achieves constant rate for long messages, where rate is measured as the ratio of ciphertext length to plaintext length. For full collusion resistance, however, the ciphertext size is linear in the number of users N. For comparison, our new system generates ciphertexts of size $O(\sqrt{N})$ and achieves constant rate (rate = 1) for long messages by using hybrid encryption (i.e. encrypting a short message-key using the traitor tracing system and encrypting the long data by using a symmetric cipher with the message-key).

Many traitor tracing systems, including ours, assume that the tracer is a trusted party and require that the tracer's key TK be kept secret. Some exceptions are [26, 27, 35, 19, 9]. Similarly, many traitor tracing systems, including ours, assume that the pirate decoder is stateless. Kiayias and Yung [18] show how to strengthen traitor tracing systems to handle stateful decoders.

Finally, we note that binary fingerprinting codes [8, 33] are closely related to traitor tracing (binary refers to the fact that the code is defined over a binary alphabet). In fact, it is known [6] that any binary fingerprinting code gives rise to a fully collusion-resistant traitor tracing system with *constant* size ciphertexts. The private key size, unfortunately, is quite large. Using [8] the private key size is $\tilde{O}(N^3)$ and using [33] it is $\tilde{O}(N^2)$.

2 Traitor Tracing and Private Linear Broadcast Encryption

In Appendix A we review the precise definition of a traitor tracing system. However, instead of directly building a traitor tracing system we build a simpler primitive called *Private Linear Broadcast Encryption* (PLBE). We first define secure PLBEs below and then briefly explain how a PLBE is used for traitor tracing. The resulting tracing algorithm makes explicit a tracing technique used in [4, 23, 18]. Then in the remainder of the paper we build a secure PLBE.

2.1 Description of Private Linear Broadcast Encryption

A PLBE is comprised of the following four algorithms:

Setup$_{\text{LBE}}(N, \lambda)$ The setup algorithm takes as input N, the number of users in the system, and the security parameter λ. The algorithm runs in polynomial time in λ and outputs a public key PK, a secret key TK, and private keys $\mathrm{K}_1, \dots, \mathrm{K}_N$, where K_u is given to user u.

Encrypt$_{\text{LBE}}(\mathrm{PK}, M)$ Takes as input a public key PK, and a message M and outputs a ciphertext C. This algorithm is used to encrypt a message to all N users.

TrEncrypt$_{\text{LBE}}(\mathrm{TK}, i, M)$ Takes as input a secret key TK, an integer i satisfying $1 \leq i \leq N + 1$, and a message M. It outputs a ciphertext C. This algorithm encrypts a message to a set $\{i, \ldots, N\}$ and is primarily used for traitor tracing. We will require below that *TrEncrypt*$_{\text{LBE}}(\mathrm{TK}, 1, M)$ outputs a distribution on ciphertexts that is indistinguishable from the distribution generated by *Encrypt*$_{\text{LBE}}(\mathrm{PK}, M)$.

Decrypt$_{\text{LBE}}(j, \mathrm{K}_j, C, \mathrm{PK})$ Takes as input a private key K_j for user j, a ciphertext C, and the public key PK. The algorithm outputs a message M or $\perp$.

The system must satisfy the following **correctness property**:
for all $i, j \in \{1, \ldots, N+1\}$, where $j \leq N$, and all messages M:

$$\text{Let} \quad \big(\mathrm{PK}, \mathrm{TK}, (\mathrm{K}_1, \ldots, \mathrm{K}_N)\big) \stackrel{\mathrm{R}}{\leftarrow} \mathit{Setup}_{\text{LBE}}(N, \lambda)$$

$$\text{and let} \quad C \stackrel{\mathrm{R}}{\leftarrow} \mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, i, M).$$

$$\text{If} \quad \mathbf{j} \geq \mathbf{i} \quad \text{then} \quad \mathit{Decrypt}_{\text{LBE}}(j, \mathrm{K}_j, C, \mathrm{PK}) = M.$$

Security. We define security of a PLBE system using three games. The first game just captures a consistency property which says that *TrEncrypt*$_{\text{LBE}}(\mathrm{TK}, 1, M)$ outputs a distribution on ciphertexts that is indistinguishable from the distribution generated by *Encrypt*$_{\text{LBE}}(\mathrm{PK}, M)$. The second game is a **message hiding game** and says that a ciphertext created using index $i = N+1$ is unreadable by anyone. The third game is an **index hiding game** and captures the intuition that a broadcast ciphertext created using index i reveals no non-trivial information about i. We will consider all these games for a fixed number of users N.

Game 1 – Indistinguishability. The first game says that the output of algorithm *TrEncrypt*$_{\text{LBE}}(\mathrm{TK}, 1, M)$ is indistinguishable from *Encrypt*$_{\text{LBE}}(\mathrm{PK}, M)$. The game proceeds as follows:

– **Setup** The challenger runs the *Setup*$_{\text{LBE}}$ algorithm and gives the adversary PK and the set of all private keys $\big\{\mathrm{K}_1, \ldots, \mathrm{K}_N\big\}$.
– **Challenge** The adversary gives the challenger a message M. The challenger flips a coin $\beta \in \{0, 1\}$ and computes

$$c \stackrel{\mathrm{R}}{\leftarrow} \begin{cases} \mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, 1, M) & \text{if } \beta = 0, \\ \mathit{Encrypt}_{\text{LBE}}(\mathrm{PK}, M) & \text{if } \beta = 1. \end{cases}$$

It gives C to the adversary.
– **Guess** The adversary returns a guess $\beta' \in \{0, 1\}$ of β.

We define the advantage of adversary $\mathcal{A}$ as $\mathsf{Adv}_{CG} = |\Pr[\beta' = \beta] - 1/2|$.

Game 2 – Message Hiding. The second game says that an adversary cannot break semantic security when encrypting using index $i = N + 1$. The game proceeds as follows:

- **Setup** The challenger runs the $Setup_{\text{LBE}}$ algorithm and gives the adversary PK and all secret keys $\{\mathrm{K}_1, \ldots, \mathrm{K}_N\}$.
- **Challenge** The adversary outputs two equal length messages M_0, M_1. The challenger flips a coin $\beta \in \{0, 1\}$ and sets $C \overset{\text{R}}{\leftarrow} TrEncrypt_{\text{LBE}}(\mathrm{TK}, N{+}1, M_\beta)$. The challenger gives C to the adversary.
- **Guess** The adversary returns a guess $\beta' \in \{0, 1\}$ of β.

We define the advantage of adversary $\mathcal{A}$ as $\mathsf{Adv}_{MH} = |\Pr[\beta' = \beta] - 1/2|$.

Game 3 – Index Hiding. The third game says that an adversary cannot distinguish between an encryption to index i and one to index $i + 1$ without the key K_i. The game takes as input a parameter $i \in \{1, \ldots, N\}$ which is given to both the challenger and the adversary. The game proceeds as follows:

- **Setup** The challenger runs the $Setup_{\text{LBE}}$ algorithm and gives the adversary PK and the set of private keys $\{\mathrm{K}_j \text{ s.t. } j \neq i\}$.
- **Challenge** The adversary outputs a message M. The challenger flips a coin $\beta \in \{0, 1\}$ and computes $C \overset{\text{R}}{\leftarrow} TrEncrypt_{\text{LBE}}(\mathrm{TK}, i + \beta, M)$. The challenger returns C to the adversary.
- **Guess** The adversary returns a guess $\beta' \in \{0, 1\}$ of β.

We define the advantage of adversary $\mathcal{A}$ as $\mathsf{Adv}_{IH}[i] = |\Pr[\beta' = \beta] - 1/2|$.

Now that the three games are established we are ready to define secure PLBE.

Definition 1. *We say that an N-user Private Linear Broadcast System (PLBE) is secure if for all polynomial time adversaries $\mathcal{A}$ we have that Adv_{CG}, and Adv_{MH}, and $\mathsf{Adv}_{IH}[i]$ for $i = 1, \ldots, N$, are negligible functions of λ.*

2.2 Reducing Traitor Tracing to PLBE

We briefly show that a secure PLBE gives a secure traitor tracing system. The complete details and proofs are given in the full version of the paper [7]. Let $\mathcal{E} = (Setup_{\text{LBE}}, Encrypt_{\text{LBE}}, TrEncrypt_{\text{LBE}}, Decrypt_{\text{LBE}})$ be a secure PLBE system. The derived traitor tracing system is defined as follows (we use the notation of Appendix A):

- *Setup* simply runs $Setup_{\text{LBE}}$ with the same parameters, and outputs PK as the public encryption key, TK as the secret tracing key, and the user keys identically to the PLBE scheme.
- *Encrypt* and *Decrypt* run algorithms $Encrypt_{\text{LBE}}$ and $Decrypt_{\text{LBE}}$ respectively with the same parameters.
- $Trace^{\mathcal{D}}(\mathrm{TK}, \epsilon)$, when called with oracle $\mathcal{D}$, and inputs TK and $\epsilon > 0$, does the following:

1. For $i = 1$ to $N + 1$, do the following:
   (a) The algorithm repeats the following $8(N \ln N)\lambda/\epsilon$ times:
      i. Sample M from the finite message space at random.
      ii. Let $C \xleftarrow{R} \mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, i, M)$.
      iii. Call oracle $\mathcal{D}$ on input C, and compare the output of $\mathcal{D}$ to M.
   (b) Let $\hat{p}_i$ be the fraction of times that $\mathcal{D}$ decrypted the ciphertexts correctly.
2. Let S be the set of all $i \in \{1, \ldots, N\}$ for which $\hat{p}_i - \hat{p}_{i+1} \geq \epsilon/(4N)$.
3. Output the set S as the set of guilty colluders.

Note that the running time of *Trace* is quadratic in N. It can be made $O(N \log N)$ using binary search instead of a linear scan.

Security. We prove that this traitor tracing scheme is secure. We argue that the system is semantically secure and provides secure tracing. Note that we did not explicitly require that a PLBE be semantically secure against a chosen plaintext attack to an outsider who possess no secret keys. Nevertheless, semantic security does follow straightforwardly from the three games used to define PLBE using a hybrid argument by means of the Index Hiding game.

We now briefly explain why traceability against arbitrary collusion follows from the security of the PLBE scheme. We show that the probability of winning the traceability game defined in Appendix A is negligible.

Let $p_i = \Pr[\mathcal{D}(\mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, i, M)) = M]$. We know that that $p_1 \geq \epsilon$ and p_{N+1} is negligible. The former follows from the fact that $\mathcal{D}$ is a useful decoder. The later follows directly from the PLBE message hiding game. Then there must exist some $j \in \{1, \ldots, N\}$ such that $p_j - p_{j+1} \geq \epsilon/(2N)$. By the Chernoff bound it follows that with overwhelming probability, $\hat{p}_j - \hat{p}_{j+1} \geq \epsilon/(4N)$. Hence, the set S output by $\mathit{Trace}^{\mathcal{D}}(S_{\mathcal{D}}, \mathrm{TK},\ \epsilon)$ is non-empty.

Using the notation of Game 2 from Appendix A, it remains to show that whenever $\hat{p}_j - \hat{p}_{j+1} > \epsilon/(4N)$ we have that $j \in T$. For such j we know, by Chernoff, that with overwhelming probability $p_j - p_{j+1} \geq \epsilon/(8N)$. Hence, $\mathcal{D}$ is able to distinguish $\mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, j, M)$ from $\mathit{TrEncrypt}_{\text{LBE}}(\mathrm{TK}, j+1, M)$ for random M. But since the PLBE is secure, the index hiding game implies that these two distributions are indistinguishable, unless one has K_j. It follows that the pirate who built $\mathcal{D}$ must have had K_j and therefore $j \in T$, as required. We give the full proof details in the full version of the paper.

3 Background and complexity assumptions

3.1 Bilinear maps

We review some general notions about bilinear maps and groups, with an emphasis on groups of *composite order* which will be used in our construction. We follow [5] in which composite order bilinear groups were first introduced.

Consider two finite cyclic groups $\mathbb{G}$ and $\mathbb{G}_T$ of same order $n = pq$, where p and q are distinct primes, and in which the respective group operation is efficiently computable and denoted multiplicatively. Assume the existence of an efficiently computable function $e : \mathbb{G} \times \mathbb{G} \to G_T$, with the following properties:

- (Bilinear) $\forall u, v \in G$, $\forall a, b \in \mathbb{Z}$, $e(u^a, v^b) = e(u, v)^{ab}$, where the product in the exponent is defined modulo n;
- (Non-degenerate) $\exists g \in G$ such that $e(g, g)$ has order n in G_T. In other words, $e(g, g)$ is a generator of G_T, whereas g generates G.

We will use the notation $\mathbb{G}_p, \mathbb{G}_q$ to denote the respective subgroups of order p and order q of $\mathbb{G}$.

We now review three assumptions we will use for proving our security. The first two assumptions are in prime order subgroups and the last two are over a composite group $\mathbb{G}$.

3.2 Decision 3-party Diffie-Hellman Assumption

The decision 3-party Diffie-Hellman problem is stated as follows. Given a group $\mathbb{G}_p$ of prime order p and random elements $g_p, A = g_p^a, B = g_p^b, C = g_p^c$ of $\mathbb{G}$ distinguish between $T = g_p^{abc}$ and $T = g_p^z$, where z is random in $\mathbb{Z}_p$.

We say that an algorithm $\mathcal{A}$ has advantage ϵ in solving the problem if

$$\left| \Pr[\mathcal{A}(g_p, g_p^a, g_p^b, g_p^c, g_p^{abc}) = 1] - \Pr[\mathcal{A}(g_p, g_p^a, g_p^b, g_p^c, g_p^z) = 1] \right| \geq \epsilon$$

The (t, ϵ)-decision 3-party Diffie-Hellman assumption (D3DH) is that no t-time adversary has advantage more than ϵ. Note that the decision 3-party Diffie-Hellman assumption implies the decision Bilinear Diffie-Hellman assumption. It also implies the standard linear assumption defined in [3].

3.3 Subgroup Decision Problem

The Subgroup Decision (SD) problem is stated as follows. Given a group $\mathbb{G}$ of composite order $n = pq$, where p, q are distinct (unknown) primes, and generators $g_p \in \mathbb{G}_p$ and $g \in \mathbb{G}$, distinguish between whether an element T is a random member of the subgroup $\mathbb{G}_p$ or a random element of the full group $\mathbb{G}$. That is distinguish whether T is a random element of $\mathbb{G}_p$ or $\mathbb{G}$.

We say that an algorithm $\mathcal{A}$ has advantage ϵ in solving the Subgroup Decision Problem if

$$\left| \Pr[\mathcal{A}(n, g_p, g, T) = 1 \ : \ T \xleftarrow{\mathrm{R}} \mathbb{G}_p] - \Pr[\mathcal{A}(n, g_p, g, T) = 1 \ : \ T \xleftarrow{\mathrm{R}} \mathbb{G}] \right| \geq \epsilon.$$

The (t, ϵ)-subgroup decision assumption is that no t-time adversary has advantage more than ϵ.

3.4 Bilinear Subgroup Decision Problem

The Bilinear Subgroup Decision (BSD) problem is stated as follows. Given a group $\mathbb{G}$ of composite order $n = pq$, where p, q are distinct (unknown) primes, and generators $g_p \in \mathbb{G}_p$ and $g_q \in \mathbb{G}_q$, distinguish a random order p element in the group $\mathbb{G}_T$ from a uniform element in the group $\mathbb{G}_T$. More precisely, we say that an algorithm $\mathcal{A}$ has advantage ϵ in solving the problem if

$$\Big| \Pr[\mathcal{A}(n, g, g_p, g_q,\ e(T, g)) = 1\ :\ T \xleftarrow{\mathrm{R}} \mathbb{G}_p] - \Pr[\mathcal{A}(n, g, g_p, g_q,\ e(T, g)) = 1\ :\ T \xleftarrow{\mathrm{R}} \mathbb{G}] \Big| \geq \epsilon.$$

The (t, ϵ)-bilinear subgroup decision assumption is that no t-time adversary has advantage more than ϵ.

4 A $\sqrt{N}$ size Private Linear Broadcast Encryption System

In this section we show how to construct a Private Linear Broadcast Encryption (PLBE) system with $O(\sqrt{N})$ size ciphertext. We can then apply the results of Section 2 and use this to build a traitor tracing scheme with $O(\sqrt{N})$ size ciphertexts.

Before we describe our construction we give some intuition as to why constructing PLBE systems with sublinear ciphertext size is difficult and describe the framework for which we will construct our PLBE system.

PLBE with Sublinear Ciphertext Size The primary difficulty in constructing a PLBE system is to provide the Index Hiding property. Using linear size ciphertexts this is easy: each user has a unique portion of the ciphertext assigned to them, which is used to encrypt the message (or session key) to just that user. If an encryptor replaces the ciphertext component of user u with a random encryption, only user u can tell the difference. All other users will be associated with a completely different portions of the ciphertext and changing u's component has no effect on their ability to decrypt.

To construct a PLBE system with sublinear size ciphertexts we must use a fundamentally different approach than the one above. Since the ciphertexts are sublinear in size, we cannot let every user have a component of the ciphertext that is dedicated for them alone. Intuitively, ciphertext components must be "shared" amongst users. Therefore, we cannot use the simple strategy of completely randomizing a portion of the ciphertext to prevent a particular user u from decrypting, since this will inherently effect the ability of other users to decrypt.

Our Framework We now give a framework for our PLBE system. We assume that the number of users, N in the system equals m^2 for some m. If the number of real users is not a square we can add "dummy" users to pad out to the next square. We arrange the users in an $m \times m$ matrix. Each user is assigned and identified by an unique tuple (x, y) where $1 \leq x, y \leq m$.

Since we will be constructing a Private Linear Broadcast Encryption system, we must have a linear ordering of the users that we can traverse. The first user in the system will be the user at matrix position $(1, 1)$ and from there we will order the users by traversing one row at a time. More precisely, the user at matrix position (x, y) will have the index $u = (x - 1)m + y$ in our ordering. We can think of this as a "row-major" ordering.

We can now refer to our Private Linear Broadcast Encryption scheme in terms of positions on the matrix. An encryption to position (i, j) means that a user at position (x, y) will be able to decrypt the message if either $x > i$ or both $x = i$ and $y \geq j$. With this notation, the Index Hiding game property states that:

- For $j < m$ it is difficult to distinguish between an encryption of a message to (i, j) from $(i, j + 1)$ without the key of user $(x = i, y = j)$.
- For $j = m$ it is difficult to distinguish an encryption of a message to position $(i, j = m)$ to that of one to $(i + 1, j = 1)$ without the key of user $(i, j = m)$.

The use of pairwise notation for referring to users and encryptions will be a purely notational convenience for describing our system.

4.1 Our Construction

Our construction makes use of bilinear maps of composite order n, where $n = pq$ and p and q are primes. In describing our scheme we will often use p or q in a subscript to denote if a group element is in the subgroup of order p or order q. *The key algebraic fact that underlies our scheme is that if g_p is any element from the order p subgroup (which we call $\mathbb{G}_p$) and g_q is any element from the order q subgroup (which we call $\mathbb{G}_q$), then we have: $e(g_p, g_q) = 1$.*

When the *TrEncrypt*$_{\text{LBE}}$ algorithm encrypts to an index (i, j) it creates ciphertext components for every column and every row. The keys of user (x, y) are structured in such a way that in order to decrypt he must pair the ciphertext components from row x, with the ciphertext components from column y. The encryption algorithm works by creating ciphertexts in the following way.

Column Ciphertext Components. (1) Ciphertexts for columns greater than or equal to j are "well formed" in both subgroups. (2) However, for a column that is less than j, the encryption algorithm will create a ciphertext that is well formed in the $\mathbb{G}_q$ subgroup, but random in the $\mathbb{G}_p$ subgroup.

Row Ciphertext Components. (1) Ciphertexts for rows less than i are completely random. Therefore, any user whose row index is less than x will not be able to decrypt. (2) The ciphertext components for row i are well formed in both subgroups. A user with row index i will be able to decrypt if his column index is greater than or equal to j. If it is less than j, the randomized ($\mathbb{G}_p$) part of the column ciphertext will scramble the result of pairing the row and column ciphertexts together. (3) Finally, for rows greater than i the ciphertext components will be well formed elements in the $\mathbb{G}_q$ subgroup only. A user with row index greater than i will be able to decrypt no matter what his column is, because the pairing will "cancel out" the randomized ($\mathbb{G}_p$) part of any column ciphertext component with the row ciphertext component that lives in $\mathbb{G}_q$.

The decryption algorithm for a user (x, y) will attempt to decrypt a ciphertext in the same manner no matter what the target index (i, j) is. The structure of the ciphertext will restrict decryption to only be successful for a user (x, y) if $x > i$ or $x = i$ and $y \geq j$. Additionally, since the attempted decryption procedure is independent of (i, j) a user can only learn whether his decryption was successful or not and the system will be private.

We describe the four algorithms that compose our PLBE system:

Setup$_{\text{LBE}}$*(*$N = m^2, 1^\kappa$*)* The setup algorithm takes as input the number of users N and a security parameter κ. It first generates an integer $n = pq$ where p, q are random primes (whose size is determined by the security parameter). The algorithm creates a bilinear group $\mathbb{G}$ of composite order n. It next creates random generators $g_p, h_p \in \mathbb{G}_p$ and $g_q, h_q \in \mathbb{G}_q$ and sets $g = g_p g_q, h = h_p h_q \in \mathbb{G}$. Next it chooses random exponents $r_1, \ldots, r_m, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_m \in \mathbb{Z}_n$ and $\beta \in \mathbb{Z}_q$.

The public key PK includes the description of the group and the following elements:

$$\begin{bmatrix} g, h, E = g^\beta, E_1 = g_q^{\beta r_1}, \ldots, E_m = g_q^{\beta r_m}, F_1 = h_q^{\beta r_1}, \ldots, F_m = h_q^{\beta r_m}, \\ G_1 = e(g_q, g_q)^{\beta \alpha_1}, \ldots, G_m = e(g_q, g_q)^{\beta \alpha_m}, H_1 = g^{c_1}, \ldots, H_m = g^{c_m} \end{bmatrix}$$

The private key for user (x, y) is generated as $K_{x,y} = g^{\alpha_x} g^{r_x c_y}$. Finally, the authority's secret key K includes factors p, q along with exponents used to generate the public key.

TrEncrypt$_{\text{LBE}}$*(*$\mathrm{K}, M, (i, j)$*)* The *TrEncrypt*$_{\text{LBE}}$ algorithm is a secret key algorithm used by the tracing authority. The algorithm encrypts a message M to the subset of receivers that have row values greater than i or both row value equal to i and column values greater than or equal to j.

The encryption algorithm will take as input the secret key, a message $M \in \mathbb{G}_T$ and an index i, j. The encryption algorithm first chooses random $t \in \mathbb{Z}_n$, $w_1, \ldots, w_m$, $s_1, \ldots, s_m \in Z_n$, $z_{p,1}, \ldots, z_{p,j-1} \in \mathbb{Z}_p$, and $(v_{1,1}, v_{1,2}, v_{1,3})$, … , $(v_{i-1,1}, v_{i-1,2}, v_{i-1,3}) \in \mathbb{Z}_n^{(3)}$.

For each row x we create four ciphertext components $(R_x, \tilde{R}_x, A_x, B_x)$ as follows:

$$\begin{array}{llll} \text{if } x > i : R_x = g_q^{s_x r_x} & \tilde{R}_x = h_q^{s_x r_x} & A_x = g_q^{s_x t} & B_x = M e(g_q, g)^{\alpha_x s_x t} \\ \text{if } x = i : R_x = g^{s_x r_x} & \tilde{R}_x = h^{s_x r_x} & A_x = g^{s_x t} & B_x = M e(g, g)^{\alpha_x s_x t} \\ \text{if } x < i : R_x = g^{v_{x,1}} & \tilde{R}_x = h^{v_{x,1}} & A_x = g^{v_{x,2}} & B_x = e(g, g)^{v_{x,3}} \end{array}$$

For each column y the algorithm creates values $C_y, \tilde{C}_y$ as:

$$\begin{array}{ll} \text{if } y \geq j : C_y = g^{c_y t} h^{w_y} & \tilde{C}_y = g^{w_y} \\ \text{if } y < j : C_y = g^{c_y t} g_p^{z_{p,y}} h^{w_y} & \tilde{C}_y = g^{w_y} \end{array}$$

Note that the ciphertext contains $5\sqrt{N}$ elements in $\mathbb{G}$ and $\sqrt{N}$ elements of $\mathbb{G}_T$.

In the above description there are three classes of rows. A row $x > i$ will have all its elements in the $\mathbb{G}_q$ subgroup, while the "target" row i will have its components in the full group $\mathbb{G}$. A row $x < i$ will essentially have its group elements randomly chosen. A column $y \geq j$ will be well formed, while a column $y < j$ will be well formed in the $\mathbb{G}_q$ subgroup, but not in the $\mathbb{G}_p$ subgroup.

Encrypt$_{\text{LBE}}$*(*PK, M*)* The *Encrypt*$_{\text{LBE}}$ algorithm is used by an encryptor to encrypt a message such that all the recipients can receive it. This algorithm is used during normal (non-tracing) operation to distribute content to all the receivers. The *Encrypt*$_{\text{LBE}}$ algorithm should produce ciphertexts that are indistinguishable from *TrEncrypt*$_{\text{LBE}}$ algorithm to the index $(1, 1)$ for the same message.

The encryption algorithm first chooses random $t \in \mathbb{Z}_n, w_1, \ldots, w_m, s_1, \ldots, s_m \in \mathbb{Z}_n$. For each row x the algorithm creates the four ciphertext components $(R_x, \tilde{R}_x, A_x, B_x)$ as follows:

$$R_x = E_x^{s_x} \qquad \tilde{R}_x = F_x^{s_x} \qquad A_x = E^{s_x t} \qquad B_x = M G_x^{s_x t}$$

For each column j the algorithm creates $C_y, \tilde{C}_y$ as:

$$C_y = H_y^t h^{w_y} \qquad \tilde{C}_y = g^{w_y}$$

Decrypt$_{\text{LBE}}$*(*$(x, y), K_{x,y}, C$*)* User (x, y) uses key $K_{x,y}$ to decrypt by computing:

$$B_x \cdot \Big(e(K_{x,y}, A_x) e(\tilde{R}_x, \tilde{C}_y) / e(R_x, C_y) \Big)^{-1}.$$

We observe that if the ciphertext was created from the tracing algorithm *TrEncrypt*$_{\text{LBE}}$with parameters (i, j) then the result is M if $x > i$ or $x = i$ and $y \geq j$. Additionally, it is easy to observe that if the ciphertext was created as *Encrypt*$_{\text{LBE}}$$(\text{PK}, M)$ then all parties can decrypt and receive M.

4.2 Discussion

Roughly, the size of the ciphertext is $5\sqrt{N}$ elements in $\mathbb{G}$ and $\sqrt{N}$ elements of $\mathbb{G}_T$. In practice, a message will be encrypted with a symmetric key cipher under a key K and our system will be used to transmit the key K to each user. We note that we can actually save in ciphertext size by converting our encryption system into a Key Encapsulation Mechanism (KEM). To do this we do not include the B_x values in the ciphertext, but instead user (x, y) can extract a key $K_x = e(K_{x,y}, A_x) e(\tilde{R}_x, \tilde{C}_y) / e(R_x, C_y)$. The extraction mechanism will actually derive $\sqrt{N}$ different keys $K_1, \ldots K_m$, so key K_x is used to encrypt K to for all users in row x. In practice this would be more space efficient than including $\sqrt{N}$ group elements of $\mathbb{G}_T$.

The $Encrypt_{\text{LBE}}$ algorithm requires $6\sqrt{N}$ exponentiations. The decryption algorithm is surprisingly efficient and simple, requiring only three pairing computations. Thus, decryption time is independent of the number of users in the system.

We constructed a (limited)[4] broadcast encryption system in which decryptors are oblivious as to which set of users the broadcast is targeted for. A set of colluding users will of course be able to learn some information about the target just by testing which one of them was able to decrypt. However, they should not learn anything more than what can naturally be inferred. The key to keeping the broadcast set private is that the decryption algorithm performs the same steps to attempt decryption no matter what the broadcast set is. In the next section we prove this intuition to be correct by showing that our scheme is secure in the Index Hiding game.

5 Security Proof

In this section we prove our Private Linear Broadcast Encryption system secure. We begin with the Index Hiding game, since the proof is the most interesting.

5.1 Proof of Security for Game 3 (Index Hiding)

For the Index Hiding game we must consider two cases. The first is when an adversary tries to distinguish between an encryption to (i, j) and an encryption to $(i, j+1)$ for $j < m$ and second for when an adversary tries to distinguish between an encryption (i, m) and one to $(i+1, 1)$.

In the first case we show that the difficulty of this game can be reduced to the 3-party Diffie-Hellman assumption, while the second case is more complicated since the structure of the row ciphertexts are changed. We handle the second case by constructing a sequence of hybrid experiments. Due to space requirements we give the proof of the lemma for the first case in the appendix and refer the reader to our full version of this paper [7] for the proofs of the other claims and lemmas.

Theorem 1. *Suppose that the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman, (t, ϵ_{BSD})-Bilinear Subgroup Decision, and (t, ϵ_{SD})-Subgroup Decision assumptions hold. Then no $\tilde{t}$-time adversary $\mathcal{A}$ can succeed in the Index-Hiding game with advantage greater than $(2+m)\epsilon_{D3DH} + \epsilon_{BSD} + \epsilon_{SD}$, where $\tilde{t} \approxeq t$.*

We first consider the case where an adversary $\mathcal{A}$ attempts to distinguish between an encryption to $(i+j)$ and $(i, j+1)$ where $j < m$. This is the case when the distinguishing game does not cross rows. We prove the following lemma in Appendix B.

[4] A Private Linear Broadcast Encryption system is restricted in the sets of users it can encrypt to — it can only encrypt to sets $\{i, \ldots, N\}$ for any i.

Lemma 1. *Suppose that the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman, assumption holds. Then no t-time adversary can distinguish between an encryption to (i, j) and $(i, j+1)$ in the Index Hiding game for $j < m$ with advantage $> \epsilon_{D3DH}$.*

We now turn to the more difficult case of when the adversary $\mathcal{A}$ chooses to distinguish between an encryption to (i, m) and one to $(i+1, 1)$ for some $1 \leq i < m$. This case becomes more complicated because the form of ciphertext rows will change. In our proofs we will refer to the rows with ciphertexts in the $\mathbb{G}_q$ subgroup as "greater than" rows and the the row with well formed ciphertexts in $\mathbb{G}$ as a "target" row. Additionally, when we say we "encrypt to column j" this means that we create ciphertexts for which C_y is well formed in the $\mathbb{G}_p$ subgroup for all $y \geq j$. We state our lemma and then prove it.

Lemma 2. *Suppose the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman, the (t, ϵ_{BSD}) -Bilinear Subgroup Decision, and the (t, ϵ_{SD})-Subgroup Decision assumptions hold. Then no $\tilde{t}$-time adversary $\mathcal{A}$ can succeed in the Index-Hiding game with advantage greater than $(2+m)\epsilon_{D3DH} + \epsilon_{BSD} + \epsilon_{SD}$, where $\tilde{t} \approxeq t$.*

We first define a sequence of hybrid experiments as follows:

- H_1: Encrypt to column m, row i is target row, i+1 is a "greater than" row.
- H_2: Encrypt to column $m+1$, row i is target row, i+1 is a "greater than" row.
- H_3: Encrypt to column $m+1$, row i is less than row, i+1 is a "greater than" row (no target row exists).
- H_4: Encrypt to column 1, row i is less than row, i+1 is "greater than" row (no target row exists).
- H_5: Encrypt to column 1, row i is less than row, i+1 is target row.

We prove our lemma by giving reductions for each consecutive pair of hybrid experiments. The proofs are given in [7].

Claim. Suppose that the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman assumption holds. Then no t-time adversary can distinguish between experiments H_1 and H_2 with advantage greater than ϵ_{D3DH} .

In both experiments we encrypt with row i as the target row and all C_y for $y < m$ random in the $\mathbb{G}_p$ subgroup. The experiment is whether an adversary can tell if the $\mathbb{G}_p$ component of C_m is well-formed without key $K_{i,m}$. This game is exactly the same as the one we proved above and thus we apply the result of Lemma 1. □

Claim. Suppose that the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman and the (t, ϵ_{BSD})-Bilinear Subgroup Decision assumptions hold. Then no t-time adversary can distinguish between experiments H_2 and H_3 with advantage greater than $2\epsilon_{D3DH} + \epsilon_{BSD}$.

Claim. Suppose that the (t, ϵ_{D3DH})-decision 3-party Diffie-Hellman assumption holds. Then no t-time adversary can distinguish between experiments H_3 and H_4 with advantage greater than $m \cdot \epsilon_{D3DH}$.

Claim. Suppose that the (t, ϵ_{SD})-Subgroup Decision assumption holds. Then no t-time adversary can distinguish between experiments H_4 and H_5 with advantage greater than ϵ_{SD} .

Lemma 2 follows by summing the maximum adversarial advantages across the hybrid experiments and Theorem 1 follows by observing that the bound of Lemma 1 is included in Lemma 2. □

5.2 Proof of Security for Game 1

Theorem 2. *Suppose the (t, ϵ_{SD}) Subgroup Decision assumption holds. Then for all messages M no t-time adversary can distinguish between a ciphertext created as Encrypt$_{\text{LBE}}$(PK, M) and one created as TrEncrypt$_{\text{LBE}}$($\mathrm{K}, M, (1,1)$) with advantage greater than ϵ_{SD}.*

This theorem follows by simply applying the same techniques as in our proof of Claim 5.1, so we omit the details. □

5.3 Proof of Security for Game 2 (Message Hiding)

Theorem 3. *All adversaries have advantage 0 in playing the Message Hiding game.*

The message hiding theorem is concerned with the adversaries advantage in winning the game when we encrypt to $(m+1, 1)$. However, this means that all rows will be completely random and independent of the messge, thus an adversary has 0 advantage. Essentially, the inability of the adversary to learn the message when he does not have any of the right keys is actually captured in our Index Hiding experiments. This final theorem shows that at the end the adversary learns now information about the ciphertext. □

6 Discussion

Our traitor tracing system has a number of possible interesting extensions for future work. In this section we discuss a few of these.

Public Traceability In our current system the tracing key, TK, is kept secret and only the authority is able to trace pirate boxes. In practice, it might be useful to have a system where the tracing key is public. For example, in a large content distribution system the capturing and tracing of pirate boxes or software will likely be done by different several agents each of which will need the tracing key. We would like our system to remain secure even if one of these agents and his tracing key is compromised.

In our $\sqrt{N}$ PLBE system the tracing algorithm would be public if a user was able to encrypt a message to an arbitrary set of indices (i, j). Then the user could simply run the tracing algorithm in the same way as the authority. In order to this we would need to give the user the capability to form C_y column ciphertext components that were well formed in its $\mathbb{G}_q$ subgroup, but not in the $\mathbb{G}_p$ subgroup. If we simply include an element of $\mathbb{G}_p$ in the public key our scheme will become insecure as an attacker could use this to determine which row index i a broadcast was intended for. Achieving public traceability would seem to require a more complex technique and possibly the use of a stronger assumption.

Stateful Receivers Like most other tracing traitor solutions our solution solves the tracing traitors problem in the stateless model, where the tracer is allowed to reset the pirate algorithm after each tracing query. However, there are some applications where we would like to consider a stronger model where a pirate box can retain state between each broadcast. In practice, a hardware pirate box might keep state and shut down if it detects that it is being traced.

Kiayias and Yung [18] showed a method which can handle stateful receivers if it were possible to embed watermarks in the distributed content and for a tracer to be able to observe these watermarks when interacting with a pirate algorithm. During *non-tracing* operation the broadcaster encrypts two copies of digital content, each of which has a different watermark embedded in, to a random (and hidden) index u. The encryption is such that all users with index less than u can decrypt the first ciphertext and all users with index greater than u can decrypt the second ciphertext. The decryption algorithm simply tries to decrypt both ciphertexts and uses whichever one results in a well-formed plaintext. The tracing algorithm will create ciphertexts in an *identical* manner to the regular encryption algorithm. The tracer will simply observe which watermarks are embedded in every probing ciphertext and use this information to identify the traitor. Since, the regular broadcast and tracing algorithms are identical a pirate box is unable to leverage its ability to maintain state.

In our current construction, our PLBE scheme is only secure if the pirate constructing the pirate decoder has not seen encryptions to arbitrary indices. However, if we were able to find a new PBLE algorithm that was secure under chosen-plaintext queries to arbitrary indicies then we could implement the techniques of Kiayias and Yung. We would simply set up two PLBE systems in which the users were given the opposite indices in each system. The user with index u in the first system has index $N+1-u$ in the second system.

7 Conclusions and Open Problems

We constructed the first fully collusion resistant traitor tracing system with sublinear size ciphertexts and constant size private keys. In particular, our system has ciphertexts of size $O(\sqrt{N})$ where N is the number of users in the system and the time for decryption is independent of N. We achieve our traitor tracing system by first introducing a simpler primitive we call private linear broadcast encryption (PLBE) that we show can give a traitor tracing system. Then, we built an efficient PLBE system by making novel use of bilinear groups of composite order.

One interesting open problem is to create a version of our traitor system that allows for public traceability. This would allow both for the tracer to be untrusted and could be used to give a solution that is secure against stateful receivers. Additionally, it is an open problem to see if one can get smaller than $\sqrt{N}$ size ciphertexts with small private keys.

References

[1] Adam Barth, Dan Boneh, and Brent Waters. Privacy in encrypted content distribution using private broadcast encryption. In *Financial Cryptography '06*, 2006.

[2] O. Berkman, M. Parnas, and J. Sgall. Efficient dynamic traitor tracing. In *Proceedings of SODA '00*, 2000.

[3] Dan Boneh, Xavier Boyen, and Hovav Shacham. Short group signatures. In Matt Franklin, editor, *Proceedings of Crypto 2004*, LNCS. Springer-Verlag, August 2004.

[4] Dan Boneh and Matthew K. Franklin. An efficient public key traitor tracing scheme. In *CRYPTO '99: Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pages 338–353, London, UK, 1999. Springer-Verlag.

[5] Dan Boneh, Eu-Jin Goh, and Kobbi Nissim. Evaluating 2-dnf formulas on ciphertexts. In Joe Kilian, editor, *Proceedings of Theory of Cryptography Conference 2005*, volume 3378 of *LNCS*, pages 325–342. Springer, 2005.

[6] Dan Boneh and Moni Naor. Tracing traitors with constant size ciphertext using binary fingerprinting codes. Unpublished, 2002.

[7] Dan Boneh, Amit Sahai, and Brent Waters. Fully collusion resistant traitor tracing with short ciphertexts and private keys. In *Eurocrypt '06*, 2006. Full version available at `http://eprint.iacr.org/2006/045`.

[8] Dan Boneh and James Shaw. Collusion secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5):1897–1905, 1998. Extended abstract in Crypto '95.

[9] Hervé Chabanne, Duong Hieu Phan, and David Pointcheval. Public traceability in traitor tracing schemes. In *EUROCRYPT*, pages 542–558, 2005.

[10] Benny Chor, Amos Fiat, and Moni Naor. Tracing traitors. In *CRYPTO '94: Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology*, pages 257–270, London, UK, 1994. Springer-Verlag.

[11] Benny Chor, Amos Fiat, Moni Naor, and Benny Pinkas. Tracing traitors. *IEEE Transactions on Information Theory*, 46(3):893–910, 2000.

[12] Yevgeniy Dodis and Nelly Fazio. Public key trace and revoke scheme secure against adaptive chosen ciphertext attack. In *Public Key Cryptography - PKC 2003*, volume 2567 of *LNCS*, pages 100–115, 2003.

[13] A. Fiat and M. Naor. Broadcast encryption. In *Proceedings of Crypto '93*, volume 773 of *LNCS*, pages 480–491. Springer-Verlag, 1993.

[14] Amos Fiat and T. Tassa. Dynamic traitor tracing. In *Proceedings of Crypto '99*, volume 1666 of *LNCS*, pages 354–371, 1999.

[15] Eli Gafni, Jessica Staddon, and Yiqun Lisa Yin. Efficient methods for integrating traceability and broadcast encryption. In *CRYPTO '99: Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pages 372–387, London, UK, 1999. Springer-Verlag.

[16] M. T. Goodrich, J. Z. Sun, , and R. Tamassia. Efficient tree-based revocation in groups of low-state devices. In *Proceedings of Crypto '04*, volume 2204 of *LNCS*, 2004.

[17] D. Halevy and A. Shamir. The lsd broadcast encryption scheme. In *Proceedings of Crypto '02*, volume 2442 of *LNCS*, pages 47–60, 2002.

[18] Aggelos Kiayias and Moti Yung. On crafty pirates and foxy tracers. In *ACM Workshop in Digital Rights Management – DRM 2001*, pages 22–39, London, UK, 2001. Springer-Verlag.

[19] Aggelos Kiayias and Moti Yung. Breaking and repairing asymmetric public-key traitor tracing. In Joan Feigenbaum, editor, *ACM Workshop in Digital Rights Management – DRM 2002*, volume 2696 of *Lecture Notes in Computer Science*, pages pp. 32–50. Springer, 2002.

[20] Aggelos Kiayias and Moti Yung. Traitor tracing with constant transmission rate. In *EUROCRYPT '02: Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques*, pages 450–465, London, UK, 2002. Springer-Verlag.

[21] K. Kurosawa and Y. Desmedt. Optimum traitor tracing and asymmetric schemes. In *Proceedings of Eurocrypt '98*, pages 145–157, 1998.

[22] Shigeo Mitsunari, Ryuichi Sakai, and Masao Kasahara. A new traitor tracing. *IEICE Trans. Fundamentals*, E85-A(2):481–484, 2002.

[23] Dalit Naor, Moni Naor, and Jeffrey B. Lotspiech. Revocation and tracing schemes for stateless receivers. In *CRYPTO '01: Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology*, pages 41–62, London, UK, 2001. Springer-Verlag.

[24] Moni Naor and Benny Pinkas. Threshold traitor tracing. In *CRYPTO '98: Proceedings of the 18th Annual International Cryptology Conference on Advances in Cryptology*, pages 502–517, London, UK, 1998. Springer-Verlag.

[25] Moni Naor and Benny Pinkas. Efficient trace and revoke schemes. In *FC '00: Proceedings of the 4th International Conference on Financial Cryptography*, pages 1–20, London, UK, 2001. Springer-Verlag.

[26] B. Pfitzmann. Trials of traced traitors. In *Proceedings of Information Hiding Workshop*, pages 49–64, 1996.

[27] B. Pfitzmann and M. Waidner. Asymmetric fingerprinting for larger collusions. In *Proceedings of the ACM Conference on Computer and Communication Security*, pages 151–160, 1997.

[28] Reihaneh Safavi-Naini and Yejing Wang. Sequential traitor tracing. In *Proceedings of Crypto '00*, volume 1880 of *LNCS*, pages 316–332, 2000.

[29] Alice Silverberg, Jessica Staddon, and Judy L. Walker. Efficient traitor tracing algorithms using list decoding. In *Proceedings of ASIACRYPT '01*, volume 2248 of *LNCS*, pages 175–192, 2001.

[30] Jessica N. Staddon, Douglas R. Stinson, and Ruizhong Wei. Combinatorial properties of frameproof and traceability codes. Cryptology ePrint 2000/004, 2000.

[31] D. Stinson and R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes. *SIAM Journal on Discrete Math*, 11(1):41–53, 1998.

[32] D. Stinson and R. Wei. Key preassigned traceability schemes for broadcast encryption. In *Proceedings of SAC '98*, volume 1556 of *LNCS*, 1998.

[33] Gabor Tardos. Optimal probabilistic fingerprint codes. In *Proceedings of STOC '03*, pages 116–125, 2003.

[34] V. To, R. Safavi-Naini, and F. Zhang. New traitor tracing schemes using bilinear map. In *Proceedings of 2003 DRM Workshop*, 2003.

[35] Yuji Watanabe, Goichiro Hanaoka, and Hideki Imai. Efficient asymmetric public-key traitor tracing without trusted agents. In *Proceedings CT-RSA '01*, volume 2020 of *LNCS*, pages 392–407, 2001.

A Definition of Tracing Traitors

Initially, we view a pirate decoder $\mathcal{D}$ as a probabilistic circuit that takes as input a ciphertext C and outputs some message M or $\perp$. A Traitor-Tracing system, then, consists of the following four algorithms:

Setup(N, λ) The setup algorithm takes as input N, the number of users in the system, and the security parameter λ. The algorithm runs in polynomial time in λ and outputs a public key BK, a secret tracing key TK, and private keys $\mathrm{K}_1, \ldots, \mathrm{K}_N$, where K_u is given to user u.

Encrypt(BK, M) Encrypts M using the public broadcasting key BK and outputs ciphertext C.

Decrypt$(j, \mathrm{K}_j, C, \mathrm{BK})$ Decrypt C using the private key K_j of user j. The algorithm outputs a message M or $\perp$.

Trace$^{\mathcal{D}}(\mathrm{TK},\ \epsilon)$ The tracing algorithm is an oracle algorithm that is given as input the tracing key TK and a parameter ϵ, and runs in time polynomial in the security parameter λ and $1/\epsilon$. Only values of ϵ that are polynomially related to λ are considered valid inputs to *Trace*. The tracing algorithm queries the pirate decoder $\mathcal{D}$ as a black-box oracle, as defined above. It outputs a set S which is a subset of $\{1, 2, \ldots, N\}$.

The system must satisfy the following **correctness property**:
for all $j \in \{1, \ldots, N\}$ and all messages M:

$$\text{Let} \quad \big(\mathrm{BK}, \mathrm{TK}, (\mathrm{K}_1, \ldots, \mathrm{K}_N)\big) \xleftarrow{\mathrm{R}} \mathit{Setup}(N, \lambda) \quad \text{and} \quad C \xleftarrow{\mathrm{R}} \mathit{Encrypt}(\mathrm{BK}, M).$$

$$\text{Then} \qquad \mathit{Decrypt}(j, \mathrm{K}_j, C, \mathrm{BK}) = M.$$

Security. We define security of the traitor tracing scheme in terms of the following two natural game.

Game 1. The first game is the standard **Semantic Security Game**. It says that the system is semantically secure to an outsider who does not possess any of the private keys. Since this is a standard notion we do not give the game details here. We define the advantage of adversary $\mathcal{A}$ in winning this game as $\mathsf{Adv}_{SS} = |\Pr[\beta' = \beta] - 1/2|$.

Game 2. The second game captures the notion of **Traceability against arbitrary collusion**. For a given N, λ and ϵ (where $\epsilon = 1/f(\lambda)$ for some polynomial f), the game proceeds as follows (both challenger and adversary are given N, λ, and ϵ as input):

1. The adversary $\mathcal{A}$ outputs a set $T = \{u_1, u_2, \ldots, u_t\} \subseteq \{1, \ldots, N\}$ of colluding users.
2. The challenger runs *Setup*(N, λ) and provides BK and $\mathrm{K}_{u_1}, \ldots, \mathrm{K}_{u_t}$ to $\mathcal{A}$. It keeps TK to itself.
3. The adversary $\mathcal{A}$ outputs a pirate decoder $\mathcal{D}$.
4. The challenger now runs $\mathit{Trace}^{\mathcal{D}}(\mathrm{TK}, \epsilon)$ to obtain a set $S \subseteq \{1, \ldots, N\}$. Note that *Trace* is only given black-box oracle access to $\mathcal{D}$.

We say that the adversary $\mathcal{A}$ wins the game if the following two conditions hold:

– The decoder $\mathcal{D}$ is useful. That is, for a randomly chosen M in the finite message space, we have that

$$\Pr[\mathcal{D}(\mathit{Encrypt}(\mathrm{BK}, M)) = M] \geq \epsilon$$

– The set S is either empty, or is not a subset of T.

We denote by Adv_{TR} the probability that adversary $\mathcal{A}$ wins this game.

Definition 2. *We say that an N-user Traitor Tracing system is secure if for all polynomial time adversaries $\mathcal{A}$ and any constant $\epsilon > 0$ we have that Adv_{MH} and Adv_{TR} are negligible functions of λ.*

We emphasize that Game 2 places no limit on the size of the coalition under the control of the adversary. Furthermore, the pirate decoder need not be perfect. It only needs to play valid content with probability ϵ. Finally, note that we are modeling a stateless (resettable) pirate decoder — the decoder is just an oracle and maintains no state between activations. Non stateless decoders were studied in [18].

In the full version of the paper we describe a more restrictive access model to the pirate decoder $\mathcal{D}$. PLBE enables tracing even in this more restrictive model.

B Proof of Lemma 1

For this distinguishing experiment we will show that distinguishing between whether an encryption is to position (i, j) or $(i, j+1)$ is as hard as the 3-party Diffie-Hellman assumption. Since, the assumption is in a prime order group the simulator can know the factorization of n, the order of the group. For this game simulator will run the core part of the simulation in the $\mathbb{G}_p$ subgroup and choose all values in the $\mathbb{G}_q$ subgroup for itself. Our formal proof follows.

Suppose there exists a t-time adversary $\mathcal{A}$ that breaks the Index Hiding game with advantage ϵ. Then we build a simulator as follows. The simulator receives the 3-party Diffie-Hellman challenge from the simulator as:

$$g_p, A = g_p^a, B = g_p^b, C = g_p^c, T.$$

The challenge will be given in the subgroup of prime order p of a composite order group $n = pq$. The simulator is given the factors p, q.

Next, the simulator runs the Init phase and receives the index (i, j) from $\mathcal{A}$. Since the game will be played in the subgroup $\mathbb{G}_p$, the simulator can choose for itself everything in the $\mathbb{G}_q$ subgroup. It chooses random generators $g_q, h_q \in \mathbb{G}_q$ and random exponents $\beta, r_{q,1}, \ldots, r_{q,m}, c_{q,1}, \ldots, c_{q,m} \in \mathbb{Z}_q$. Additionally, it chooses the exponents $\alpha_1, \ldots, \alpha_m \in \mathbb{Z}_n$. It then sets $h_p = B$ and picks blinding factors $r'_{p,1}, \ldots, r'_{p,m}, c'_{p,1}, \ldots, c'_{p,m} \in \mathbb{Z}_p$.

The simulator is now able to create the public and secret keys as follows. It first publishes $g = g_q g_p$ and $h = h_q B$. It creates the public keys:

$$E = g_q^{\beta} \qquad E_x = g_q^{\beta r_{q,x}} \qquad F_x = h_q^{\beta r_{q,x}}$$
$$G_x = e(g_q, g_q)^{\beta \alpha_x} \qquad H_y = \begin{cases} g_q^{c_{q,y}} g_p^{c'_{p,y}} & : y \neq j \\ g_q^{c_{q,y}} C^{c'_{p,y}} & : y = j \end{cases}$$

Next, it creates the private keys for all users except (i, j) as:

$$K_{x,y} = \begin{cases} g^{\alpha_x} g_q^{r_{q,x} c_{q,y}} g_p^{r'_{p,x} c'_{p,y}} & : x \neq i, y \neq j \\ g^{\alpha_x} g_q^{r_{q,x} c_{q,y}} B^{r'_{p,x} c'_{p,y}} & : x = i, y \neq j \\ g^{\alpha_x} g_q^{r_{q,x} c_{q,y}} C^{r'_{p,x} c'_{p,y}} & : x \neq i, y = j \end{cases}$$

We note that all the simulator creates public and private with the same distribution as the real scheme.

In the challenge phase the adversary first gives the simulator a message $M \in \mathbb{G}_T$. The simulator then chooses exponents $(v_{1,1}, v_{1,2}, v_{1,3}), \ldots, (v_{i-1,1}, v_{i-1,2}, v_{i-1,3}\ ) \in \mathbb{Z}_n^{(3)}$, and exponents $s_{q,i}, \ldots, s_{q,m} \in \mathbb{Z}_q$ and $t_q \in \mathbb{Z}_q$. Additionally, it chooses random $s'_p \in \mathbb{Z}_p$, $z_{p,1}, \ldots, z_{p,j-1} \in \mathbb{Z}_p$, $w'_1, \ldots, w'_m \in \mathbb{Z}_n$.

It then creates the ciphertext as:

$$\begin{array}{llll}
\text{if } x > i: & R_x = g_q^{s_{q,x} r_{q,x}} & \tilde{R}_x = h_q^{s_{q,x} r_{q,x}} \\
& A_x = g_q^{s_{q,x} t_q} & B_x = M e(g_q, g_q)^{\alpha_x s_{q,x} t_q} \\
\text{if } x = i: & R_x = g_q^{s_{q,x} r_{q,x}} g_p^{s'_p r'_{p,x}} & \tilde{R}_x = h_q^{s_{q,x} r_{q,x}} B^{s'_p r'_{p,x}} \\
& A_x = g^{s_{q,x} t_q} A^{s'_p} & B_x = M e(g_q, g_q)^{\alpha_x s_{q,x} t_{q,x}} e(g_p, A)^{\alpha_x s'_p} \\
\text{if } x < i: & R_x = g^{v_{x,1}} & \tilde{R}_x = h^{v_{x,1}} \\
& A_x = g^{v_{x,2}} & B_x = e(g, g)^{v_{x,3}} \\
\text{if } y > j: & C_y = g_q^{c_{q,y} t_q} h^{w'_y} & \tilde{C}_y = A^{-c'_{p,y}} g^{w'_y} \\
\text{if } y = j: & C_y = g_q^{c_{q,y} t_q} T h^{w'_y} & \tilde{C}_y = g^{w'_y} \\
\text{if } y < j: & C_y = g_q^{c_{q,y} t_q} g_p^{z_{p,y}} h^{w'_y} & \tilde{C}_y = g^{w'_y}
\end{array}$$

If T forms a 3-party Diffie-Hellman tuple then the ciphertext is a well-formed encryption to the indices (i, j), otherwise if T is randomly chosen it is a encryption to $(i, j + 1)$. The simulator will receive a guess γ from $\mathcal{A}$ and it will simply repeat this guess as its answer to the 3-party Diffie-Hellman game. The simulator's advantage in the Index Hiding game will be exactly equal to $\mathcal{A}$'s advantage. $\square$