Leakage Resilience of the Blom’s Key Distribution Scheme

Michał Jastrzębski** and Stefan Dziembowski***

Institute of Informatics, University of Warsaw

Abstract. We initiate the study of the leakage-resilience of the information-theoretic key distribution schemes. Such schemes, originally proposed in the 1980s, have recently attracted a lot of interest in the systems community. This is because, due to their extreme efficiency, they can be executed on low-cost devices such as sensors, where the use of the public-key cryptography is infeasible. We argue that the study of leakage resilience of such schemes is particularly well-motivated, since, unlike more expensive devices, the sensors (or other similar devices) are unlikely to be physically resilient to leakage.

We concentrate on the classical scheme of Blom (CRYPTO 1982), since it is known to be optimal in a large class of such schemes. We model the leakage as an input-shrinking function. In this settings we show that Blom’s scheme is leakage-resilient in a very strong model, where the adversary can (1) compromise completely some nodes in a “standard” way, and (2) leak information jointly from the remaining nodes. The amount leakage that we can tolerate can be up to \((0.5 - \epsilon)\) of the total amount of information on the leaking nodes. We also show that this bound is optimal, by providing an attack that breaks the scheme if more leakage is available to the adversary. This attack works even in a weaker model, where the nodes leak information independently.

In the proof we make use of the theory of the randomness extractors. In particular we use the fact that inner product over a finite field is a good 2-source extractor. This is possible since the Blom’s scheme is based on the matrix multiplication.

1 Introduction

A recent trend in theoretical cryptography, initiated by [34, 29, 28], is to design schemes that are provably-secure even if implemented on devices
that are not fully trusted. The motivation for this research comes from the fact that, instead of breaking the mathematical foundations of a cryptosystem, in real life it is often much easier to attack its physical implementation. Such “physical attacks” are usually based on the side-channel information about the internals of the cryptographic device that the adversary can obtain by measuring its running-time, electromagnetic radiation, power consumption (see e.g. [36]), or even by actively tampering with it (see e.g. [3]) in order to force it to reveal information about its secrets. Practitioners have developed several remedies to these attacks, however, they are usually ad-hoc and lack a formal security argument.

Contrary to the approach taken by the practitioners, security of the constructions developed by the theoreticians is always analyzed formally in a well-defined mathematical model, and hence covers a broad class of attacks, including those that are not yet known, but may potentially be invented in the future. Over the last few years several models for passive and active physical attacks have been proposed and schemes secure in these models have been constructed (see e.g. [29, 24, 28, 19, 38, 1, 14, 31, 35, 20, 22, 7, 13, 22, 30, 10, 18, 6, 23, 26, 7]). Some of these papers [29, 28, 22, 30, 18, 26] present the so-called general compilers i.e. algorithms that transform any cryptographic functionality into a “physically-secure” one. These generic constructions, although very inspiring theoretically, are of a limited practical relevance mostly because of the huge blow-up in the complexity of the computed functionality. In another class of papers the authors develop new schemes for concrete cryptographic task such as stream-ciphers [19, 38], public-key primitives [31, 35, 7, 13, 10, 32], multi-party computation protocols [6] or zero-knowledge schemes [23]. While some of these schemes can be quite efficient, it is unclear if they will ever be used in practice, one of the reasons being that their deployment would require a change in the existing industrial standards.

Therefore an alternative natural approach is to analyze the leakage-resilience of existing cryptosystems in order to find among them those that exhibit good leakage-resilience properties. One example of such work is the influential paper of Akavia et al. [1] where the authors show leakage-resilience of the public-key encryption scheme of Regev [40] and the identity-based encryption scheme of Gentry et al. [25]. Another example can be found in [2] where the leakage-resilience of the Okamoto identification scheme [37] is shown and used in a construction of a signature scheme secure in the bounded-retrieval model. The schemes in these examples are computationally secure and we are not aware of any non-trivial exam-
ple of a practical information-theoretically secure scheme whose leakage-resilience has been shown in the literature.

Usually the information-theoretically secure schemes are considered not very practical since, for various reasons, they are cumbersome to use in real-life, the classical example being the one time pad encryption scheme that requires the users to store very large keys. Nevertheless, some of the information-theoretically secure schemes have found practical applications, due to their simplicity and efficiency. One of such examples is the Shamir’s secret sharing scheme [42], which is used as a building block for several cryptographic protocols. Another prominent example are the information-theoretically secure key distribution schemes (KDS). The leakage resilience of one of such schemes, namely the classical Blom’s scheme [4] is the main topic of this paper. Below we first give a brief introduction to the KDS’s and then informally describe our leakage model.

**Key Distribution Schemes (KDS)** Since currently the main application of the KDS’s are the sensor networks we will use the sensor network terminology. We note, however, that the mathematical idea appeared much before the emergence of such networks (Blom’s paper was published in 1982), and such schemes have also other applications, one example being the HDCP (High-Bandwidth Digital Content Protection) system created by Intel.

A sensor network consists of a large number $m$ of autonomous devices denoted by natural numbers $1, \ldots, m$. In many cases, sensors are being deployed in hostile environment and are exposed to diverse malicious adversaries. In this case security of communication between nodes becomes essential. Sensors need to be able to communicate directly with each other over encrypted and authenticated channels. Such communication requires that each pair of nodes share a common secret key (which is used by this pair for encryption). After they are deployed, the nodes should communicate without relying on any trusted third party. Therefore, before the deployment, a trusted setup phase, called the key predistribution, is executed. Since the sensors have a limited computing power the use of the public-key cryptography is not an option, and the solutions based on the symmetric-key primitives are preferred.

One obvious solution is for each sensor to store the secret key to pair with any other node. In this case the common key of a pair would be secret and known only to the nodes in this pair. This approach is not practical because of the nature of sensors - the devices have limited memory, whereas the total number of stored keys grows quadratically in the
size of the network. Another extreme is to give the same fixed key to every sensor in the network. While this scheme would be very memory efficient its security is quite low, since the adversary who compromises just one node, and extracts the key from it, would be able to decrypt the communication between each pair of the not compromised nodes sharing the same key.

Rolf Blom in [4] proposed a key distribution scheme which provides a nice tradeoff between security and memory efficiency. Description of the Blom key distribution scheme as well as its security can be found in Section 2. Informally speaking, the idea of [4] is to fix a number $n$ of nodes that the adversary needs to compromise in order to break the security of the system. More precisely: as long as the number of compromised nodes is smaller than $n$ then the keys shared by any pair of not compromised nodes is unknown to the adversary. The size of the stored information on each sensor is $n \cdot |K|$ bits (where $|K|$ is the length of the key that the sensors establish). Note that Blom’s scheme can be viewed as a generalization of the schemes described above: by setting $n = m$ we obtain the scheme with high resilience against compromising of the nodes, but high memory requirements, and by setting $n = 1$ we obtain the other extreme case (low resilience and low memory requirements). For a formal analysis of Blom scheme see e.g. [5], where it is also shown that the Blom’s scheme is optimal in terms of the amount of data that the sensors need to store (as a function of $n, m$ and $|K|$). Practical key predistribution schemes were also proposed by Eschenauer and Gligor [21] who construct a scheme where two nodes can establish the common key only with some probability. Other key predistribution schemes providing different tradeoffs can be found for instance in [8, 16].

**Leakage resilience of the KDS’s** The sensors that execute the key distribution schemes are supposed to be low-cost, and therefore they cannot be assumed to be leakage-proof, as physical protections against leakages is expensive. Therefore analyzing the side-channel leakage-resilience of the key distribution schemes is particularly well-motivated, and it is natural to pick the Blom’s scheme as the first target for this analysis (for its optimality and simplicity). Let us start with the description of our leakage model. As we already mentioned, several models have been proposed for reasoning about the side-channel leakages, for example the early approaches considered the leakages of the individual bits only [12, 29]. In this work we follow a very popular paradigm in which the leakage is modeled as an *input shrinking function*, i.e. a function $f$ whose output is much
shorter than its input (the length of the output of \( f \) will be called the \textit{amount of leakage}). Such functions were first proposed in cryptography in the so-called bounded-storage model of Maurer [33]. Later, they were used to define the memory leakage occurring during the virus attacks in the bounded-retrieval model [17, 11, 2]. In the context of the side-channel leakages they were first used in [19] with an additional restriction that the memory is divided into two separate parts that do not leak information simultaneously, and in the full generality in the paper of Akavia et al. [1].

\textbf{Our contribution} In this paper we use the model of Akavia et al. i.e. we do not impose any restrictions on the input of the leakage function, except that we require only that the size of its output is bounded. Several other papers in this model, already mentioned in the introduction, have been published in recent years. A popular tool that is used in these works are the randomness extractors (see e.g. [41]). Since some of these extractors are based on linear algebra (in particular the famous inner product extractor of Chor and Goldreich [9]) hence the Blom’s scheme, which itself uses the matrix multiplication, seems to be a promising candidate for a leakage-resilient key distribution scheme. We confirm this intuition in this paper.

More precisely, but still informally, we show that the Blom’s scheme is leakage-resilient in the following sense. In our model we allow the adversary to both perform the standard “non-leakage” attacks (i.e. to compromise the nodes) and to leak from the uncompromised nodes\(^1\). Recall that \( n - 1 \) is the maximal number of nodes that the adversary can compromise while attacking the scheme without leakage. In order to show the leakage-resilience we will now treat \( n \) as a security parameter and allow the adversary to compromise significantly less nodes than \( n - 1 \).

Let \( j \leq n - 1 \) denote the number of nodes that the adversary actually compromised. It turns out (cf. Theorem 2) that if \( j \) is close to \( n - 1 \) than the leakage-resilience of the Blom’s scheme is very low. In particular if \( j = n - 1 \) then the scheme can be broken with leaking just \( |K| \) bits from the uncompromised nodes, and in general Blom’s scheme can be broken in with leakage of size that is \( \approx 0.5(n - j)^2 \cdot |K| \). Therefore we assume that \( n - j \) is linear in \( n \). Under this assumption we show (Theorem 1) that the key established between a pair of nodes, 1 and 2, say, remains secret even if the adversary (additionally to the information that he got by compromising the nodes) learns the value of the leakage function \( f \) on the concatenation of the internal data of all the other nodes. The maximal

\(^1\) For simplicity of the notation in our formal model the leakage function is in fact also applied to the compromised nodes.
amount of leakage (i.e. the length of the output of \( f \)) that we can tolerate is at most \( c(n - j)^2 \cdot |K| \), where \( c \) can be any constant such that \( c < 0.5 \). A small caveat is that the leakage function cannot depend on the identifiers\(^2\) of the sensors 1 and 2, and hence in this sense it is non-adaptive.

Traditionally, the leakage resilience of the cryptographic schemes is measured in terms of the relative leakage \( \lambda \) that they can tolerate, which is defined as the ratio between the length of the output and the input of the leakage function. In order to talk in these terms let us assume that the parameter \( n \) is linear in \( m \), i.e. there exists a constant \( \alpha \) such that \( n = \alpha m \). Our Theorem 1 implies (cf. Corollary 1) that the maximal achievable value of \( \lambda \) depends on the fraction \( \gamma \) of compromised nodes in the following way: \( \lambda \leq c \cdot (\alpha - \gamma)^2 / \alpha \) (where \( c \) is any constant such that \( c < 0.5 \)). Hence, e.g., if the adversary did not compromise any node then \( \lambda \) can be close to 0.5, if we choose \( \alpha \) close to 1.

As highlighted above we also prove that the bound given in Theorem 1 is optimal by showing an attack that uses leakage of size \( \approx 0.5(n - j)^2 \cdot |K| \).

This attack actually works in a weaker model, where the sensors leak information independently, i.e. separate leakage function is applied to each sensor and the restriction on the leakage size concerns the sum of the lengths of the outputs of these functions (cf. Definition 3). The leakage functions in this attack can be chosen in advance, however they need to adaptively depend on the identifiers of the sensors, hence we call it an adaptive model. In Section 8 we show an even weaker attack (with slightly worse bounds) when the adversary does not even need to adaptive in the sense that the leakage functions do not depend on the identifiers of the sensors.

Before we present our contribution in detail we first describe formally (in Section 2) the Blom scheme and its security in the standard model without leakages. Then, in Section 3 we incorporate leakage into this model. Our main results are stated in Section 4 and their proofs are given in Sections 5-6.

## 2 Blom’s key distribution scheme

In this section we give a formal definition of the Blom’s scheme that was already informally introduced in Section 1. Let \( \mathbb{F} \) denote some finite field (for instance \( GF(p) \)), \( M_n(\mathbb{F}) \) and \( S_n(\mathbb{F}) \) will denote respectively the set of all matrices and symmetric matrices of size \( n \times n \) over field \( \mathbb{F} \). If \( m \) is

\(^2\) In the Blom’s scheme every node \( i \) has its identifier \( i \) that is chosen randomly and is used by the other nodes to compute the keys for communicating with \( i \).
the size of the network (number of sensors participating in the protocol), then the scheme will be parametrized by $n$ being a security parameter indicating the number of nodes which needs to get compromised before the adversary is able to break the scheme. The protocol setup requires that all secrets are being predistributed before the network is deployed. To formalize, the protocol setup for an $(\mathbb{F}, n, m)$-Blom scheme (for the network size $m$ and security parameter $n$) works as follows

- Central server selects publicly known identifiers $x_1, \ldots, x_m$, where each $x_i \in \mathbb{F}^n$. Nodes are labeled 1, 2, $\ldots$, $m$ so that $x_i$ is the identifier of the node $i$. Next, server chooses uniformly at random as symmetric matrix $R \in \mathbb{S}_n(\mathbb{F})$. The matrix is kept secret and will be destroyed after this step. The secret stored on device $i$ consists of value $Rx_i \in \mathbb{F}^n$. After this step each node is associated with two values $(x_i, Rx_i)$ where $x_i$ is publicly known and $Rx_i$ is kept secret. The network is deployed.

- If nodes $i$ and $j$ wish to establish a shared key, then $i$ computes $x_j^T(Rx_i)$, whereas $j$ computes $x_i^T(Rx_j)$.

The symmetry of $R$ assures that

$$x_i^T R x_j = (x_i^T R x_j)^T = x_j^T R^T x_i = x_j^T R x_i$$

which guarantees that computed keys match on both sides of the channel.

It can be proven that if every $n$ of the identifiers $x_1, \ldots, x_m$ are linearly independent, then the adversary compromising any $n - 1$ of the nodes (apart from $i$ and $j$ of course), cannot determine the secret key which would be computed by nodes $i$ and $j$. Moreover, such an adversary cannot gain any knowledge on this key. More formally, we can show [4] that

$$H(x_i^T R x_j | \text{secrets of } n - 1 \text{ nodes}) = H(x_i^T R x_j)$$

where $H(\cdot)$ and $H(\cdot | \cdot)$ denote respectively the entropy and conditional entropy of a random variable. Formal proofs and extensions to conference keys can be found, for instance, in [5]. In practical applications, to ensure this notion of security, identifiers can be chosen as columns from a Vandermonde matrix (for definitions, consult for instance [27]). In this paper we will assume that the identifiers are simply uniformly and independently at random from $\mathbb{F}^n$. This, of course, can generate problems if, by accident, the selected identifiers are not linearly independent. Fortunately, the probability that this happens will be negligible and will disappear under the asymptotic notation.
3 Leakage attacks on the Blom’s scheme

Consider a regular Blom scheme described in the previous chapter which is determined by $F$, $n$ and $m$ being respectively the finite field, dimension of the random symmetric matrix and the number of participants. Matrix $R$ is randomly chosen from $S_n(F)$. Participants are receiving publicly known random identifiers $x_1, \ldots, x_m \in F^n$ as well as secrets of the form $Rx_i$. The adversary wants to find the common key of two nodes. To simplify the exposition let us assume that these nodes are always 1 and 2, and therefore the key between them is $x_1^T Rx_2 = x_2^T Rx_1$.

In our work, we will consider the scheme in a general framework of the memory leakage. We allow the adversary to choose any function $f$ and apply it to the secrets $Rx_3, \ldots, Rx_m$ giving $f(Rx_3, \ldots, Rx_m)$. The function $f$ models an arbitrary memory leakage or an eavesdropping device which can be used on stored secrets. It is worth noticing that compromising several nodes is just a simple sub-case of the leakage function. Our results show that if the size of the image of $f$ is small enough, the distribution of $x_1^T Rx_2$ conditioned on the value of $f$ is close to the uniform.

We wish to make our adversary adaptive, which means that it would choose the leakage function $f$ based on the publicly known identifiers. Note that it is impossible to consider the leakage model when an adversary would be able to choose $f$ based on all the identifiers, namely $x_1, \ldots, x_m$. Indeed, in this case, if $m > n + 2$, $f$ could just compute $R$ based on $Rx_3, \ldots, Rx_m$ and output $x_2^T Rx_2$. Size of the image of $f$ equals $|F|$ and completely compromises the protocol, which is unacceptable. Instead, we will allow an adversary to choose $f$ based on $x_3, \ldots, x_m$.

A part from the model in which the function $f$ operates simultaneously on $Rx_3, \ldots, Rx_m$ we would consider the leakage model in which the adversary is choosing functions $f_3, \ldots, f_m$ operating respectively on $Rx_3, \ldots, Rx_m$. It seems that such an adversary would be significantly weaker than the one using the joint function. Our results show that it is not the case. We will also consider the non-adaptive adversary which needs to choose $f$ before identifiers are being set up.

Our security models are incorporating both the notion of regular sensor compromising (as considered in the usual security analysis of the Blom scheme) as well as leakage functions. Next sections provide detailed descriptions of our security definitions in terms of games between an adversary $A$ and an oracle $\Omega$. 
**Notation** For a random variable $X$ we would define its support as $\text{supp}(X) := \{x : \mathbb{P}(X = x) > 0\}$. We will also need a notion of min-entropy $H_\infty(X) := \min_{x \in \text{supp}(X)} -\log(\mathbb{P}(X = x))$. Statistical distance is defined as $\Delta(X;Y) := \frac{1}{2} \sum_x |\mathbb{P}(X = x) - \mathbb{P}(Y = x)|$. By a distance to uniform distribution, conditioned on a random variable we will understand $d(X|Y) := \sum_x \mathbb{P}(Y = x) \cdot d(X|Y = x)$ where $d(X) := \Delta(X;U)$ for $U$ uniform independent on $X$.

**Strong adaptive adversary (joint leakage)** Fix some parameters $n, m, j, k$ and the finite field $\mathbb{F}$. There are $m$ participants in the protocol. Consider the following game between the oracle $\Omega$ and the adversary $A$.

1. $A$: The adversary chooses $j$ nodes among $3, 4, \ldots, m$ which we will call compromised. We may assume that the nodes chosen by an adversary are numbered $3, 4, \ldots, j + 2$.
2. $\Omega$: The oracle chooses $R \in \mathbb{S}_n(\mathbb{F})$ uniformly at random. Oracle chooses $x_1, \ldots, x_m$ from $\mathbb{F}^n$ independently uniformly at random. Values of $x_3, \ldots, x_m$ are sent to $A$.
3. $A$: The adversary chooses a function $f : \mathbb{F}^{n(m-2)} \rightarrow \{1, \ldots, |\mathbb{F}|^k\}$
4. $\Omega$: The oracle sends to the adversary $x_1, x_2$ as well as $Rx_3, \ldots, Rx_{j+2}$ (secrets from compromised nodes) and $f(Rx_3, \ldots, Rx_m)$

Observe that, in order to simplify the notation, we measure the size of the leakage not in terms of bits but in terms of the field elements and the function $f$ can also be viewed as having a type $f : \mathbb{F}^{n(m-2)} \rightarrow \{0, 1\}^{k \log_2 |\mathbb{F}|}$. For an adversary $A$, by $\text{View}^A$ we will denote the vector of values of all random variables which were observed by the adversary during its game with the oracle

**Definition 1.** We say that the $(\mathbb{F}, n, m)$-Blom scheme is strongly $(j, k, \epsilon)$-secure if for any Adversary $A$ in the game above we have that

$$d(x_1^T R x_2 | \text{View}^A) \leq \epsilon.$$  

**Weak adaptive adversary (separate leakages)** The setting is the same as in the strong adversary model. The difference between this model and the strong one is in Steps 3 and 4:

3. $A$: Adversary chooses a functions $f_3, \ldots, f_m$ such that $f_i : \mathbb{F}^n \rightarrow \{1, 2, \ldots, |\mathbb{F}|^{k_i}\}$ where $k_3 + \cdots + k_m \leq k$
4. \( \Omega \): Oracle sends to the adversary \( x_1, x_2 \) as well as \( Rx_3, \ldots, Rx_{j+2} \) (secrets from compromised nodes) and \( f_3(Rx_3), \ldots, f_m(Rx_m) \)

**Definition 2.** We say that \((\mathbb{F}, n, m)\)-Blom scheme described above is weakly \((j, k, \epsilon)\)-secure if for any Adversary \( A \) in the game above we have that

\[
d(x_1^T Rx_2 | \text{View}^A) \leq \epsilon.
\]

4 Our results

In this section we state our main security results.

**Theorem 1.** For every \( n \) consider a \((\mathbb{F}, n, m)\)-Blom scheme and \( j(n) \) and \( k(n) \) such that \( n - j(n) = \Omega(n) \). Let \( c < 0.5 \) be an arbitrary constant. If \( k(n) \leq c(n - j(n))^2 \) then the scheme is strongly \((j(n), k(n), |\mathbb{F}|^{-\Omega(n)})\)-secure. Moreover the constant hidden in \( \Omega(n) \) does not depend on \( |\mathbb{F}| \).

The proof of this theorem is presented in Section 6 and the main technical machinery is developed in Section 5. Observe that the leakage resilience (measured by the parameters \( j(n), k(n) \) and \( |\mathbb{F}|^{-\Omega(n)} \)) does not depend on the number of the parties, but only on the difference between the parameters \( n \) and \( j(n) \). Traditionally, the leakage resilience of a scheme is measured in terms of the relative size \( \lambda \) of the leakage with respect to the total size of data that can leak, which in our case is \( n(m - 2) \). The following corollary of Theorem 1 serves for interpreting our result in this way. We will assume that there exists a constant \( \alpha \) such that \( n = \alpha m \), in other words: the security parameter (and hence the size of the data stored by each node) is linear in the number of sensors.

**Corollary 1.** Let \( \lambda, \gamma, \alpha \in [0, 1] \) and \( c < 0.5 \) be constants such that

\[
\lambda \leq \frac{c \cdot (\alpha - \gamma)^2}{\alpha},
\]

and let \( \mathbb{F} \) be an arbitrary finite field. Then for every \( m \) the \((\mathbb{F}, n, m)\)-Blom scheme (with \( n = \alpha m \)) is \((\gamma \cdot m, \lambda \cdot n(m - 2), |\mathbb{F}|^{-\Omega(n)})\)-secure.

**Proof.** Let \( j := \gamma \cdot m \) and \( k = \lambda \cdot n(m - 2) \). It is a straightforward calculation that \( k \leq c(n - j(n))^2 \). Hence, by Theorem 1, the corollary is true. \( \square \)

The corollary implies that, as long as the number \( j \) of compromised nodes is a constant fraction of the total number of nodes, we can tolerate a constant relative leakage \( \lambda \) with respect to the total size of the data.
Observe that in the extreme case when no parties are compromised we have \( \gamma = 0 \), and hence \( \lambda \) is at most \( c_\alpha \approx \alpha/2 \), which, for \( \alpha \) close to 1, means that the relative leakage can be close to 0.5.

As mentioned in the introduction, the parameters obtained in Theorem 1 are optimal. Indeed, even if we pass to the weak model, the adversary can fully compromise the protocol for \( k(n) \approx |F|^{0.5(n-j(n))^2} \). To formalize, we can prove the following.

**Theorem 2.** For every \( n \) consider a \((F,n,m)\)-Blom scheme and \( n \) and \( k(n) \) such that \( k(n) = 0.5(n - j(n))(n - j(n) + 1) \). Such scheme is not weakly \((j(n),k(n),\epsilon)\)-secure for any \( \epsilon < 0.1 \).

The proof appears in the full version of this paper [?]. Theorems 1 and 2 show together, that two considered adaptive models, namely models with joint and separate leakages can be considered as equivalent in terms of asymptotic security. This means that allowing the adversary to compute leakage function "mixing" the secrets stored at different nodes, gives him no significant advantage over the model in which we allow only to leakages from separate nodes. This may be viewed as counterintuitive, as one may expect that joint leakage model would allow to significantly reduce the leakage size necessary to compromise the protocol. The main technical tool that we use to prove Theorem 1 is the lemma that appears in the next section.

### 5 The main technical lemma

**Lemma 1.** Let \( X, Y \) be independent random vectors uniformly distributed over \( F^n \). If \( R \) is a random matrix uniformly distributed among \( \text{supp}(R) \subset M_n(F) \) and independent of \((X,Y)\), then for an arbitrary function \( f : M_n(F) \rightarrow \{1,\ldots,|F|^k\} \) we have that

\[
d(X^T R Y | X, Y, f(R)) \leq |F|^{u-n+1} + \frac{|F|^{k/8n}}{|\text{supp}(R)|^{1/8n}} |F|^{(n-u+7)/2}
\]

for any \( u < n \)

One may be tempted to think about this fact as a simple corollary from leftover hash lemma (LHL), because we can treat the variable \( X^T R Y \) as coming from a family of hash functions \( H_{(x,y)}(R) = x^T R y \) indexed by random choice of \((x,y)\). It turns out that this is not the case since this hash family is only about \( 2/|F| \) - almost universal which does not allow us to use the LHL. Trying modifications of the LHL proof using, for instance,
conditioning on the rank of the random matrix $R$ also does not seem to provide sufficiently good estimates.

Before we present our proof we need some technical tools. We start with the useful notion of an inner product extractor, then develop linear algebra lemmas and finally move towards the core of the proof.

5.1 Strong two source extractors [9]

Definition 3. We will call a function $\text{Ext} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ a strong $(k_1, k_2, \epsilon)$-two-source extractor if and only if for any independent random variables $X, Y$ in $\mathbb{F}^n$ such that $H_\infty(X) \geq k_1$ and $H_\infty(Y) \geq k_2$ we have

$$d(\text{Ext}(X,Y) | X) \leq \epsilon.$$ 

If $X$ and $Y$ are random vectors in $\mathbb{F}^n$, then recall that by $X^T Y$ we will denote the regular dot product of vectors $X$ and $Y$. For $\text{Ext}(X,Y) = X^T Y$, [15] provides a simple proof of the bound on $\epsilon$ for $|\mathbb{F}| = 2$. This result can be extended to an arbitrary finite field $\mathbb{F}$ and follows easily from the work of [39].

Theorem 3 ([39]). (Inner product extractor) The function $\text{Ext} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ defined as $\text{Ext}(X,Y) = X^T Y$ is a strong $(k_X, k_Y, |\mathbb{F}|(2^{n-k_X} + |\mathbb{F}|(n+1)/2 - (u+k_Y)/2))$-two source extractor for any $u \leq k_X$.

5.2 Linear algebra tools

This section develops linear algebra tools ending with a following combinatorial lemma which is crucial for the proof of our results.

Lemma 2. Let $\{M_v\}_{v \in \mathbb{F}^n}$ be any family of subsets of $M_n(\mathbb{F})$ satisfying

$$\forall v \in \mathbb{F}^n \forall M_1, M_2 \in M_v M_1 v = M_2 v.$$ 

Then

$$\sum_{v \in \mathbb{F}^n} |M_v| \leq |\mathbb{F}|^4 \cdot |\mathbb{F}|^n \cdot \left| \bigcup_{v \in \mathbb{F}^n} M_v \right|^{1-1/4n}.$$ 

Lemma 2 shows that sets $M_v, v \in \mathbb{F}^n$ must be overlapping considerably compared to $|\bigcup M_v|$. Technical proof of this lemma appears in the full version of this paper [?].
5.3 Proof of Lemma 1

Let us recall that we have set an integer $n$ and an arbitrary function $f : \mathbb{M}_n(\mathbb{F}) \to \{1, 2, \ldots, |\mathbb{F}|^k\}$ for some $k$. We are given independent random variables $X, Y, R$ such that $X, Y$ are uniformly random vectors from $\mathbb{F}^n$ and $R$ is random matrix chosen uniformly at random from $\text{supp}(R) \subset \mathbb{M}_n(\mathbb{F})$. Let us define

$$\mathcal{M}_i = \{ M \in \text{supp}(R) : f(M) = i \}.$$

and note that $\sum_i |\mathcal{M}_i| = |\text{supp}(R)|$. Also, let $|R| := |\text{supp}(R)|$. We wish to estimate

$$d(X^T R Y | X, Y, f(R))$$

$$= \sum_i \left( \mathbb{P}(R \in \mathcal{M}_i) \sum_{y \in \mathbb{F}^n} \mathbb{P}(Y = y) d \left( X^T R y \mid \left| X, Y = y, R \in \mathcal{M}_i \right. \right) \right).$$

For a fixed $y$ and $\mathcal{M}_i$ we may look at

$$d \left( X^T R y \mid X, Y = y, R \in \mathcal{M}_i \right)$$

as an inner product extraction from independent random variables $X$ and $Ry$, conditioned on event that $\{ R \in \mathcal{M}_i, Y = y \}$. By $H^i_{\infty}(y)$ we will denote

$$H^i_{\infty}(y) := H_{\infty}(Ry | R \in \mathcal{M}_i).$$

Of course $H_{\infty}(X) = n \log(|\mathbb{F}|)$. Let us fix some $u < n \log(|\mathbb{F}|)$. Using independency of $X, Y, R$, Theorem 3 concerning inner product extraction gives us

$$d \left( X^T R y \mid X, Y = y, R \in \mathcal{M}_i \right) \leq |\mathbb{F}| \left( \frac{2^u}{|\mathbb{F}|^n} + |\mathbb{F}|^{(n+1)/2} 2^{-(u + H^i_{\infty}(y))/2} \right).$$
Plugging this estimation into \((*)\) gives us
\[
(*) \leq |F| \cdot \frac{2^n}{|F|^n} + \sum_i \left( \mathbb{P}(R \in M_i) \sum_{y \in F^n} \mathbb{P}(Y = y) \mathbb{P}^{(n+3)/2 - \left(u + H_{\infty}^{i}(y)\right)/2} \right)
\]
\[
\leq |F| \cdot \frac{2^n}{|F|^n} + \frac{|F|^{-n}}{|R|} \sum_i \left| M_i \right| \sum_{y \in F^n} \mathbb{P}^{(n+3)/2 - \left(u + H_{\infty}^{i}(y)\right)/2}
\]
\[
= |F| \cdot \frac{2^n}{|F|^n} + \frac{|F|^{-n}}{|R|} \mathbb{P}^{(n+3)/2 - u/2} \sum_i \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2}
\]
\[
= |F| \cdot \frac{2^n}{|F|^n} + \frac{|F|^{-n}}{|R|} \mathbb{P}^{(n+3)/2 - u/2} \sum_i \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2}
\]
(2)

Let us focus on estimating the term \(\sum_i \left( \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2} \right)\). We will use the following Holder inequality
\[
\sum_i \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2} \leq |F| \cdot \frac{2^n}{|F|^n} \sum_i \left( \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2} \right)
\]
(3)

valid for any \(0 < \alpha < 1\) and \(x_i > 0\). Definition of \(H_{\infty}^{i}(y)\) and Lemma 2 yield
\[
\sum_i \left( \left| M_i \right| \sum_{y \in F^n} 2^{-H_{\infty}^{i}(y)/2} \right)
\]
\[
= \sum_i \left( \left| M_i \right| \sum_{y \in F^n} \sqrt{\max_{z \in F^n} \mathbb{P}(Ry = z | R \in M_i)} \right)
\]
\[
= \sum_i \left( \sqrt{|M_i|} \sum_{y \in F^n} \sqrt{\max_{z \in F^n} \{|R : Ry = z\} \cap M_i|} \right)
\]
\[
\leq |F|^{n/2} \sum_i \left( \sqrt{|M_i|} \cdot \sqrt{\sum_{y \in F^n} \max_{z \in F^n} \{|R : Ry = z\} \cap M_i|} \right)
\]
(4)
\[
\leq |F|^{n/2} \sum_i \left( |M_i| \right)^{1-1/8n}
\]
(5)
\[
\leq |F|^{n/2} |F|^{k/8n} |R|^{1-1/8n}
\]
(6)

Inequalities (5) and (7) come from applying (4) for \(\alpha = 1/2\) and \(\alpha = 1 - 1/8n\) respectively. Estimation in (6) can be deduced by applying Lemma 2
to receive that
\[ \sum_{y \in \mathbb{F}^n} \max_{z \in \mathbb{F}^n} \{|R : Ry = z\} \cap M_i| \leq |\mathbb{F}|^4 |\mathbb{F}|^n |M_i|^{1-1/4n} \]

Plugging obtained inequality (7) into (3) we receive

\[
d(X^T RY | X, Y, f(R)) \leq |\mathbb{F}|^{2n} + |\mathbb{F}|^{-n} \frac{|R|^{(3n+7)/2} - u/2 |\mathbb{F}|^{k/8} |R|^{1-1/8n}}{|\text{supp}(R)|^{1/8n}}
\]

for any \( u \leq n \log |\mathbb{F}| \), which completes the proof with substitution \( u := u/\log |\mathbb{F}| \).

\[\square\]

### 6 Proof of Theorem 1

Theorem 1 can be seen as a corollary from Lemma 1. We will need a following lemma, counting the number of symmetric matrices given its values on a chosen set of vectors.

**Lemma 3.** Let \( c_1, c_2, \ldots, c_j \) and \( v_1, v_2, \ldots, v_j \) be vectors such that \( c_i \in \mathbb{F}^n \), \( v_i \in \mathbb{F}^n \) and \( j \leq n \). In this setting, either

\[ \{M \in S_n(\mathbb{F}) : \forall i Mv_i = c_i\} = \emptyset \]

or

\[ |\{M \in S_n(\mathbb{F}) : \forall i Mv_i = c_i\}| \geq |\mathbb{F}|^{(n-j)(n-j+1)/2} \]

**Proof.** The size of such set is minimal for \( v_1, \ldots, v_j \) being linearly independent which we will assume from this point. We are able to choose vectors \( v_{j+1}, \ldots, v_n \) so that \( v_1, \ldots, v_n \) form a basis of \( \mathbb{F}^n \). To uniquely determine \( M \) it is enough to set \( Mv_{j+1}, \ldots, Mv_n \). Note, that

\[ M \in S_n(\mathbb{F}) \iff \forall a, b : v_a^T M v_b = v_b^T M v_a. \]

This means, that if there exist \( a, b \leq j \) such that \( v_a^T c_b \neq v_b^T c_a \) then we fall into the first case of our lemma. Otherwise, \( Mv_{j+1} \) must fulfill \( v_a^T c_{j+1} = v_{j+1}^T c_a \) for all \( a < j + 1 \), which gives us a linear space of dimension \( n - j \). Similarly, having set \( c_{j+1} \), the value \( c_{j+2} := Mv_{j+2} \) can be set into one of \( |\mathbb{F}|^{n-j-1} \) possibilities, etc. In total we receive \( |\mathbb{F}|^{(n-j)+(n-j-1)+\ldots+1} \) possibilities of choosing \( M \) which completes the proof.

\[\square\]
Proof (of Theorem 1). Proving the Theorem 1 is equivalent to estimating
\[
  d \left( \left| X_1^T R X_2 \right| \left| X_1, \ldots, X_m, R X_3, \ldots, R X_{j+2}, f(X_3, \ldots, X_m, R X_3, \ldots, R X_m) \right| \right)
\]
where \( X_1, \ldots, X_m \) are \( m \) independent random vectors from \( \mathbb{F}^n \), \( R \) is a random matrix from \( S_n(\mathbb{F}) \) independent on \( (X_1, \ldots, X_m) \) and \( f \) has a type \( \mathbb{F}^n(m-2)+n(m-2) \rightarrow \mathbb{F}^k \). Denote (7) with \( D \).

\( X_1, \ldots, X_m \) are random variables denoting identifiers attached to participants. Values of \( R X_3, \ldots, R X_{j+2} \) come from fully compromised nodes. One may easily observe that the statistical distance described above indeed corresponds to our model, as allowing the adversary to adapt to \( X_3, \ldots, X_m \) is equivalent to increasing the number of arguments of function \( f \).

Assume that we have set \( m-2 \) vectors \( x_3, x_4, \ldots, x_m, x_i \in \mathbb{F}^n \). To shorten notation, \( R(c_3, \ldots, c_{j+2}) \) will denote the event \( \{ R x_3 = c_3, \ldots, R x_{j+2} = c_{j+2} \} \). Define \( D(x_3, \ldots, x_m) \) to be equal to
\[
  d \left( \left| X_1^T R X_2 \right| \left| X_1, X_2, R x_3, \ldots, R x_{j+2}, f(x_3, \ldots, x_m, R x_3, \ldots, R x_m) \right| \right).
\]

Obviously, this is equal to
\[
  \sum_{(c_3, \ldots, c_{j+2}) \in \mathbb{F}^n} \mathbb{P}(R(c_3, \ldots, c_{j+2})) \cdot d \left( \left| X_1^T R X_2 \right| \left| X_1, X_2, f(x_3, \ldots, x_m, R x_3, \ldots, R x_m), R(c_3, \ldots, c_{j+2}) \right| \right)
\]
Lemma 6 implies, that for any \( c_3, \ldots, c_{j+2}, \) we receive
\[
  |\{ a : \mathbb{P}(R = a | R(c_3, \ldots, c_{j+2})) > 0 \}| = 0
\]
or
\[
  |\{ a : \mathbb{P}(R = a | R(c_3, \ldots, c_{j+2})) > 0 \}| \geq |\mathbb{F}|^{(n-j)(n-j+1)/2}.
\]
Also, computing \( f(x_3, \ldots, x_m, R x_3, \ldots, R x_m) \) is equivalent to computing \( g(R) \) for some \( g : \mathcal{M}_n(\mathbb{F}) \rightarrow \mathbb{F}^k \). Using Lemma 1 we obtain
\[
  D(x_3, \ldots, x_m) \leq |\mathbb{F}|^{u-n+1} + |\mathbb{F}|^{k/8n-(n-j)(n-j+1)/16n}|\mathbb{F}|^{(n-u+7)/2}.
\]
Averaging over \( x_3, \ldots, x_m \) results in the same estimation for \( D \):
\[
  D \leq |\mathbb{F}|^{u-n+1} + |\mathbb{F}|^{k/8n-(n-j)(n-j+1)/16n}|\mathbb{F}|^{(n-u+7)/2}.
\]
Plugging $k = c(n - j)^2$ leads us to

$$D \leq |F|^{u-n+1} + |F|^{7/2} \exp \left\{ \log(|F|) \left( \frac{n-u}{2} + \frac{(2c-1)(n-j)^2}{16n} \right) \right\}$$

As $n - j(n) = \Omega(n)$, there is a constant $j_0$ such that $n - j(n) \geq j_0 n$, which gives

$$D \leq |F|^{u-n+1} + |F|^{7/2} \exp \left\{ \log(|F|) \left( \frac{n-u}{2} - \frac{(1-2c)j_0^2 n}{16} \right) \right\}$$

By an arbitrary choice of $u < n$ we obtain $D = |F|^{-\Omega(n)}$.

\[\square\]

7 An even weaker adversary

As described in Section 1 we can weaken our adversary even more (with respect to the one in the proof on Theorem 2), and the negative result will still hold (with slightly worse parameters). The definition of this new weaker model follows.

Weak non-adaptive adversary (separate leakages) The setting and the security definition is the same as in the weak adaptive model (cf. Definition 3). This only difference is that in this case the adversary will not adapt to $x_3, \ldots, x_m$. More precisely, the security game is as follows:

1. $\mathcal{A}$: Adversary chooses $j$ nodes among $3, 4, \ldots, m$ which we will call compromised. We may assume that the nodes chosen by an adversary are numbered $3, 4, \ldots, j + 2$. Adversary chooses functions $f_3, \ldots, f_m$ such that $f_i : \mathbb{F}^n \to \{1, 2, \ldots, |\mathbb{F}|^{k_i}\}$ where $k_3 + \cdots + k_m \leq k$
2. $\mathcal{O}$: Oracle chooses $R \in S_n(\mathbb{F})$ uniformly at random. Oracle chooses $x_1, \ldots, x_m$ from $\mathbb{F}^n$ independently uniformly at random. Oracle sends to the adversary $x_1, x_2, x_3, \ldots, x_m$ as well as $R x_3, \ldots, R x_{j+2}$ (secrets from compromised nodes) and $f_3(R x_3), \ldots, f_m(R x_m)$

Definition 4. We say that $(\mathbb{F}, n, m)$ scheme described above is weakly non-adaptively $(j, k, \epsilon)$-secure if for any Adversary $\mathcal{A}$ in the game above we have

$$d(x_1^T R x_2 | \text{View}^\mathcal{A}) \leq \epsilon.$$ 

Theorem 4. For every $n$ consider a $(\mathbb{F}, n, m)$-Blom and $j(n)$ and $k(n)$ such that $k(n) = 0.5(n - j(n))(n - j(n) + 1)$. Such scheme is is not weakly non-adaptively $(j(n), k(n), \epsilon)$-secure for any $\epsilon < 0.5(1 - 1/|\mathbb{F}|)^{n-1}$.
Proof of this theorem can be found in a full version of this paper [7].

In fact, in proofs of Theorem 2 as well as Theorem 4 we are explicitly constructing adversaries breaking the scheme with probabilities as indicated in statements. Obtained result would suggest that the scheme in the non-adaptive model cannot be yet considered as compromised, as 

\[(1 - 1/|F|)^{n-1}\]

for growing \(n\) tends exponentially to 0. Note, however, that this term is strongly dependent on \(|F|\) (as we already mentioned, this is not the case in Theorem 1). In fact, for practical consideration, if we would take \(|F| \approx 2 \cdot 10^9\), (32-bit integers) then even for very large \(n \approx 2 \cdot 10^9\) we have

\[(1 - 1/|F|)^{n-1} \approx 1/e\]

which allows us to treat this error as constant for practical applications. Thus, we may say that the leakage of size \(k(n) \approx 0.5(n - j(n))^2\) compromises the protocol in practical setting and demonstrates that in practice non-adaptive and adaptive models are also equivalent in terms of security.

References


