Amplifying Privacy in Privacy Amplification

Divesh Aggarwal
New York University

Yevgeniy Dodis*
New York University

Zahra Jafargholi†
Northeastern University

Eric Miles†
Northeastern University

Leonid Reyzin
Boston University

Abstract

We study the classical problem of privacy amplification, where two parties Alice and Bob share a weak secret X of min-entropy k, and wish to agree on secret key R of length m over a public communication channel completely controlled by a computationally unbounded attacker Eve.

Despite being extensively studied in the literature, the problem of designing "optimal" efficient privacy amplification protocols is still open, because there are several optimization goals. The first of them is (1) minimizing the *entropy loss* $L = k - m$ (it is known that the optimal value for $L = O(\lambda)$, where $\varepsilon = 2^{-\lambda}$ is the desired security of the protocol). Other important considerations include (2) minimizing the number of communication rounds, (3) maintaining security even after the secret key is used (this is called *post-application robustness*), and (4) ensuring that the protocol P does not leak some "useful information" about the source X (this is called *source privacy*). Additionally, when dealing with a very long source X, as happens in the so-called Bounded Retrieval Model (BRM), extracting as long a key as possible is no longer the goal. Instead, the goals are (5) to touch as little of X as possible (for efficiency), and (6) to be able to run the protocol many times on the same X, extracting multiple secure keys.

Achieving goals (1)-(4) (or (2)-(6) in BRM) simultaneously has remained open, and, indeed, all known protocols fail to achieve at least two of them. In this work we improve upon the current state-of-the-art, by designing a variety of new privacy amplification protocols, in several cases achieving *optimal parameters for the first time.* Moreover, in most cases we do it by giving relatively *general transformations* which convert a given protocol P into a "better" protocol P'. In particular, as special cases of these transformations (applied to best known prior protocols), we achieve the following privacy amplification protocols for the first time:

- 4-round (resp. 2-round) *source-private* protocol with *optimal entropy loss* $L = O(\lambda)$, whenever $k = \Omega(\lambda^2)$ (resp. $k > \frac{n}{2}(1-\alpha)$ for some universal constant $\alpha > 0$). Best previous constant round source-private protocols achieved $L = \Omega(\lambda^2)$.
- 3-round *post-application-robust* protocols with *optimal entropy loss* $L = O(\lambda)$, whenever $k = \Omega(\lambda^2)$ or $k > \frac{n}{2}(1-\alpha)$ (the latter is also *source-private*). Best previous post-application robust protocols achieved $L = \Omega(\lambda^2)$.
- The first BRM protocol capable of extracting the optimal number $\Theta(k/\lambda)$ of session keys, improving upon the previously best bound $\Theta(k/\lambda^2)$. (Additionally, our BRM protocol is post-application-robust, takes 2 rounds, and can be made source-private by increasing the number of rounds to 4.)

*Supported by NSF CNS Grants 1314568, 1319051, 1065288, 1017471, and Faculty Awards from Google and VMware.

†Supported by NSF grants CCF-0845003 and CCF-1319206.

1 Introduction

We study the classical problem of *privacy amplification* [BBR88, Mau92, BBCM95, MW97] (PA), in which two parties, Alice and Bob, share a weak secret X of min-entropy k and wish to agree on a close-to-uniform secret key R of length m. We consider the active-adversary case, in which the communication channel between Alice and Bob can be not only observed, but also fully controlled, by a computationally unbounded attacker Eve. The most natural quantity to optimize here is the *entropy loss* $L = k - m$ (for a given security level $\varepsilon = 2^{-\lambda}$), but several other parameters (described below) are important as well.

Aside from being clean and elegant, this problem arises in a number of applications, such as biometric authentication, leakage-resilient cryptography, and quantum cryptography. Additionally, the mathematical tools used to solve this problem (such as randomness extractors [NZ96]) have found many other applications in other areas of cryptography and complexity theory. Not surprisingly, PA has been extensively studied in the literature, as we survey below.

In the easier "passive adversary" setting (in which Eve can observe, but not modify, the communication channel), PA can be solved by applying a (strong) *randomness extractor* [NZ96], which uses a uniformly random nonsecret seed S to extract nearly uniform secret randomness from the weak secret X. A randomness extractor accomplishes passive-adversary PA in one message: Alice sends the seed S to Bob, and both parties compute the extracted key $R = \mathsf{Ext}(X; S)$. Moreover, it is known that the optimal entropy loss of randomness extractors is $L = \Theta(\log(1/\varepsilon))$ [RTS00], and this bound can be easily achieved (e.g. using the Leftover Hash Lemma [HILL99]).

Active Eve Setting: Number of Rounds vs. Entropy Loss. The situation is more complex in the "active Eve" setting. Existing one-message solutions [MW97, DKK+12] work for min-entropy $k > n/2$ and require large entropy loss $L > n - k$. It was shown by [DS02, DW09] that $k > n/2$ is necessary, and that the large entropy loss of $n - k$ is likely necessary, as well. Thus, we turn to protocols of two rounds or more rounds.

Two rounds were shown to be sufficient by [DW09], who proved, nonconstructively, the existence of two-round PA protocols with optimal entropy loss $L = \Theta(\log(1/\varepsilon))$ for any k. (This was done using a strengthening of extractors, called *non-malleable extractors*, whose existence was shown in [DW09].) Constructively, no such protocols are known, and all known constructive results sacrifice either the number of rounds, or the entropy loss, or the minimum entropy requirement. A protocol of [Li12b, Theorem 1.9] (building on [RW03, KR09, CKOR10]) sacrifices the number of rounds: it achieves $L = O(\log(1/\varepsilon))$, but only in $O(1 + \log(1/\varepsilon)/\sqrt{k})$ rounds. The protocol of [Li12b, Theorem 1.6] (building on [DW09]) sacrifices the minimum entropy requirement: it achieves $L = O(\log(1/\varepsilon))$ in two rounds, but only when $k = \Omega(\log^2(1/\varepsilon))$. Protocols of [DLWZ11, CRS12, Li12a, Li12c] make an incomparable minimum entropy requirement: they also achieve $L = O(\log(1/\varepsilon))$ in two rounds, but require that $k > n/2$ (with the exception of [Li12c], who slightly relaxed it to $k > \frac{n}{2}(1-\alpha)$ for some tiny but positive constant α). These protocols also built the first constructive non-malleable extractors when $k > n/2$. The result of [Li12b, Theorem 1.8] (building on [DLWZ11, Li12a]) further relaxes the entropy requirement to $k > \delta n$ for any constant $\delta > 0$. It also achieves $L = O(\log(1/\varepsilon))$ in two rounds, but the constant hidden in the O-notation is $g(\delta) = 2^{(1/\delta)^c}$ for some astronomical (and not even exactly known) constant c.[1] More generally, since some of the protocols mentioned above hide relatively large (or, as in the last example, even astronomical) constant factors, simpler protocols (such as [DW09] or [KR09]) may outperform asymptotically optimal ones for many realistic settings of parameters.

[1] The value c depends on some existential results in additive combinatorics. However, it appears safe to conclude that it is astronomical, which translates into "triply astronomical" $g(\delta) = 2^{(1/\delta)^c}$, even for $\delta = 0.49$.

To summarize, the landscape of existing PA protocols is rather complex, even if we consider only the tradeoff between the min-entropy, the entropy loss, and the number of rounds. The situation becomes even more complex, if one adds additional highly desirable properties: *source privacy*, *post-application robustness*, and *local computability*. We consider those next.

SOURCE PRIVACY. Intuitively, this property demands that the transcript of the protocol (even together with the derived key R!) does not reveal any "useful information" about the source X; or, equivalently (as shown by [DS05]), that the transcript does not reveal any information at all about the *distribution* of X (beyond a lower bound k on its min-entropy). For the case of passive Eve, source privacy was considered by Dodis and Smith [DS05], who showed that randomness extractors are indeed source-private. For active Eve, the only work that considered this notion is the elegant paper [BF11], which constructed a 4-round private protocol with entropy loss $L = O(\log^2(1/\varepsilon))$. Thus, unlike for PA protocols without source privacy,

(A) *no source-private PA protocol is known which achieves either optimal entropy loss* $L = O(\log(1/\varepsilon))$*, or fewer than four rounds.*

POST-APPLICATION ROBUSTNESS. Informally, the basic authenticity notion of PA protocols, called *pre-application robustness* by [DKK+12], simply states that Eve cannot force Alice and Bob to agree on different keys $R_A \neq R_B$. While easy to define, this property is likely insufficient for most applications of PA protocols, because in any two-party protocol, one party (say, Bob) has to finish before the other party. In this case, Bob is not sure if Alice ever received his last message, and must somehow decide to use his derived key R_B. In doing so, he might leak some partial information about R_B (possibly all of it!), and Eve might now use this partial (or full) information to modify the last message that Bob originally sent to Alice. Motivated by these considerations, [DKK+12] defined a strong property called *post-application robustness*, which (intuitively) requires that Eve cannot modify Bob's last message and cause Alice to output $R_A \neq R_B$, even if given Bob's key R_B.

The only protocols known to achieve post-application robustness are in [DKK+12, DW09, DLWZ11]; in particular, none of them achieves optimal entropy loss. (Recall, the protocol of [DLWZ11] for entropy $k > \delta n$ achieves entropy loss $O((1/\delta)^c \log(1/\varepsilon))$ in $O((1/\delta)^c)$ rounds for some astronomical constant c mentioned in Footnote 1.) Most protocols in [RW03, DKK+12, DW09, CKOR10, DLWZ11, CRS12, Li12a, Li12c, Li12b] are proven only for pre-application robustness (some works simply ignored the distinction). In particular,

(B) *no post-application robust, constant-round protocol with optimal entropy loss is known (with the exception of protocol of [DLWZ11] using astronomical constants mentioned above).*

LOCAL COMPUTABILITY AND REUSABILITY. Local computability is of interest when the length and the min-entropy of the source X is much larger than the desired number of extracted bits m. In such a case, it is desirable to compute the output without having to read all of the source. This property is traditionally associated with the Bounded Retrieval Model (BRM) [Dzi06, CLW06], where the random source X is made *intentionally huge*, so that X still has a lot of entropy k even after the attacker ("virus") managed to download a big fraction of X over time. For historical reasons, we will also use the term "BRM", but point out that local computability seems natural in any scenario where $k \gg m$, and not just the BRM application.

The right way to think about entropy loss in such a scenario is not via the formula $L = k - m$, because entropy from X is not "lost": much entropy remains in X even after the protocol execution, because most of X is not even accessed. In fact, the PA protocol may be run multiple times on the same X, to obtain multiple keys, until the entropy of X is exhausted. Specifically focusing on

$m = \Theta(\log(1/\varepsilon))$ (so that the extracted key can be used to achieve ε security), "optimal" reusability means the ability to extract $\Theta(k/\log(1/\varepsilon))$ keys (assuming the entropy rate of X is constant).

In the passive adversary case, optimal reusability is achievable with locally computable randomness extractors [Lu04, Vad04]. In the active adversary case, however, the story is again more complicated. The only prior work to consider local computability in this setting is the work of [DW09]. Reusability has not been explicitly considered before, but it is easy to see that the locally computable protocol of [DW09] allows the extraction of $\Theta(k/\log^2(1/\varepsilon))$ keys. Thus,

(C) *no prior locally computable protocol achieves optimal reusability.*

1.1 Our Results

In this work, we solve open problems (A), (B), and (C), by designing several new techniques for building PA protocols. Many of our techniques are *general transformations* that convert a given protocol P into a "better" protocol P'. Given a wide variety of incomparable existing PA protocols (surveyed above), this modular approach will often allow us to obtain several improved protocols in "one go".

Two Methods of Adding Source Privacy. Our first method (Section 3.2) maintains the number of rounds at 2, at the expense of using a strengthening of non-malleable extractors [DW09] (which we call *adaptive non-malleable extractors*) to derive a one-time pad to mask the "non-private" message which should be sent in the second round. (Given that we already use non-malleable extractors however, we might as well combine our protocol with the non-private protocol of [DW09] based on non-malleable extractors with similar parameters; this is what we do to keep things simple.) Our second method (Section 3.3), inspired by the specific protocol of [BF11], turns certain 2-round non-private protocols into 4-round private protocols, using standard extractors and XOR-universal hash functions. (The concrete protocol of [BF11] implicitly applied a very particular variant of our transformation to the two-round protocol of [DW09], but we get improved results using "newer" protocol [Li12b].) In particular, either one of these transformations will provide (with different tradeoffs) a positive answer to Open Question (A). For completeness, we also observe (Section 3.1) that the 1-round PA protocols of [DKK+12] are already source-private.

Pre- to Post-Application Robustness. We make a very simple transformation which converts pre-application robust protocols to post-application robust protocols, at the cost of one extra round, but with almost no increase in the entropy loss. Although very simple, it immediately gives a variety of answers to Open Question (B) (and can also be combined with our first transformation, since it preserves source privacy).

Overall, by applying our transformations above to different protocols and in various orders, we get several improvements to existing protocols, summarized in Table 1 (which includes various solutions to Questions (A), (B), and more).

Achieving Local Computability and Optimal Reusability. While only the work of [DW09] explicitly considered local computability, it is reasonable to ask if other existing protocols can be modified to be locally computable and reusable. To achieve optimal reusability, we focus on protocols with optimal entropy loss, because they have the property that the protocol transcript reduces the entropy of X by $O(\log(1/\varepsilon))$, leaving *residual* entropy of X high. They can be modified to extract a short key of length $\Theta(\log(1/\varepsilon))$, which will give optimal reusability.

To achieve local computability, extractors used within a protocol can be replaced with locally computable extractors. Indeed, the protocol of [CKOR10] seems to amenable to such modification.

[2] We observe in this paper that this protocol is private.

Result	Entropy	Rounds	Entropy Loss		Source Privacy
			Pre-app	Post-app	
[DW09] (non-expl.)	$k = \Omega(\log(1/\varepsilon))$	2	$\Theta(\log(1/\varepsilon))$	$\Theta(\log(1/\varepsilon))$	NO
This work (non-expl.)	$k = \Omega(\log(1/\varepsilon))$	2	$\Theta(\log(1/\varepsilon))$	$\Theta(\log(1/\varepsilon))$	**YES**
[DKK+12]	$k > \frac{n}{2}$	1	$n - k - \Theta(\log(1/\varepsilon))$	$\frac{n}{2} + \Theta(\log(1/\varepsilon))$	**YES**[2]
[Li12b]	$k = \Omega(\log^2(1/\varepsilon))$	2	$\Theta(\log(1/\varepsilon))$	$\Theta(\log^2(1/\varepsilon))$	NO
This work	$k = \Omega(\log^2(1/\varepsilon))$	3	$\Theta(\log(1/\varepsilon))$	$\Theta(\log(1/\varepsilon))$	NO
[BF11]	$k = \Omega(\log^2(1/\varepsilon))$	4	$\Theta(\log^2(1/\varepsilon))$	$\Theta(\log^2(1/\varepsilon))$	YES
This work	$k = \Omega(\log^2(1/\varepsilon))$	4	$\Theta(\log(1/\varepsilon))$	$\Theta(\log^2(1/\varepsilon))$	YES
This work	$k = \Omega(\log^2(1/\varepsilon))$	5	$\Theta(\log(1/\varepsilon))$	$\Theta(\log(1/\varepsilon))$	YES
[Li12c]	$k > \frac{n}{2}(1-\alpha)$	2	$\Theta(\log(1/\varepsilon))$	$\frac{n}{2}(1-\alpha) + \Theta(\log(1/\varepsilon))$	NO
This work	$k > \frac{n}{2}(1-\alpha)$	2	$\Theta(\log(1/\varepsilon))$	$\frac{n}{2}(1-\alpha) + \Theta(\log(1/\varepsilon))$	**YES**
This work	$k > \frac{n}{2}(1-\alpha)$	3	$\Theta(\log(1/\varepsilon))$	$\Theta(\log(1/\varepsilon))$	**YES**

Table 1: Our improvement (also marked in **RED**) over prior PA protocols.

However, it is not constant-round. Most other constant-round protocols with optimal entropy loss [DLWZ11, CRS12, Li12a, Li12c] use non-malleable extractors, and this approach fails, because no locally computable (even non-constructive!) instantiations of non-malleable extractors are known.

However, we observe that the 2-round, optimal entropy loss protocol of [Li12b, Theorem 1.6] does not use non-malleable extractors. Moreover, by making all extractors in that protocol locally computable, we get a locally computable, 2-round protocol. However, the security analysis of [Li12b] uses a very delicate and interdependent setting of various parameters for the security proof to go through. Hence, it is not immediately clear if this intricate proof will go though if one uses locally computable extractors. Instead, we will develop a different, *modular* analysis underlying the key ideas of [Li12b], which will give us a rigorous 2-round solution to open problem (C), as well as have other benefits we describe shortly. Specifically, we show a general transformation that turns certain (post-application) secure 2-round protocols into 2-round protocols *with optimal entropy loss* $L = O(\log(1/\varepsilon))$ *and residual min-entropy* $k' = k - O(\log(1/\varepsilon))$ (Section 5). The transformation uses two-source extractor of [Raz05] to compress the second message of the protocol to only $O(\log(1/\varepsilon))$ bits. By applying this transformation to the original (non-BRM) protocol of [DW09], we get a protocol very similar to the protocol of [Li12b], but with a much more modular and easier-to-follow security analysis. On the other hand, by using the locally computable protocol of [DW09] instead (see Section 6), we get a 2-round locally computable protocol with optimal residual entropy (and, thus, reusability), solving open problem (C).[3] Furthermore, we can add source privacy by using our 2-to-4-round transformation mentioned earlier, which can be done via local computation as well.

These results are summarized in Table 2.

Improving Entropy Loss of Post-Application Robust Protocols. As another advantage

[3]Interestingly, the main limitation of the non-BRM protocol of [Li12b] — high min-entropy requirement $k = \Omega((\log(1/\varepsilon))^2)$ — is not an issue in the BRM model. Thus, we can view our result as finding a "practical application scenario" for the very elegant communication reduction technique developed by [Li12b].

Result	Rounds	Residual Min-entropy	# Keys Extracted	Source Privacy
[DW09]	2	$k - \Theta(\log^2(1/\varepsilon))$	$\Theta(k/\log^2(1/\varepsilon))$	NO
This work	2	$k - \Theta(\log(1/\varepsilon))$	$\Theta(k/\log(1/\varepsilon))$	NO
This work	4	$k - \Theta(\log(1/\varepsilon))$	$\Theta(k/\log(1/\varepsilon))$	**YES**

Table 2: Protocols in the Bounded Retrieval Model; each extracts $\Theta(\log(1/\varepsilon))$ bits per key, is post-application robust, and requires $k = \Omega(\log^2(1/\varepsilon))$. Entries in **RED** mark our improvements.

of our modular approach, we note that the transformation described in the previous paragraph is interesting not only in the context of local computability. It also allows one to turn *post-application* robust 2-round protocols with *sub-optimal* entropy loss L into 2-round *pre-application* robust protocols with *optimal* entropy loss, which then (using our pre-application to post-application transformation described above) can be turned into 3-round *post-application* robust protocols with *optimal* entropy loss. Namely, we can obtain optimal entropy loss at the expense of one extra round. (For the BRM setting, no extra round is needed, as we only extract "short" keys of length $O(\log(1/\varepsilon))$.)

2 Preliminaries

For a set S, we let U_S denote the uniform distribution over S. For an integer $m \in \mathbb{N}$, we let U_m denote the uniform distribution over $\{0,1\}^m$, the bit-strings of length m. For a distribution or random variable X we write $x \leftarrow X$ to denote the operation of sampling a random x according to X. For a set S, we write $s \leftarrow S$ as shorthand for $s \leftarrow U_S$.

Entropy and Statistical Distance. The *min-entropy* of a random variable X is defined as $\mathbf{H}_\infty(X) \stackrel{\text{def}}{=} -\log(\max_x \Pr[X = x])$. We say that X is an (n,k)-*source* if $X \in \{0,1\}^n$ and $\mathbf{H}_\infty(X) \geqslant k$. For $X \in \{0,1\}^n$, we define the *entropy rate* of X to be $\mathbf{H}_\infty(X)/n$. We also define *average (aka conditional) min-entropy* of a random variable X conditioned on another random variable Z as

$$\mathbf{H}_\infty(X|Z) \stackrel{\text{def}}{=} -\log\left(\mathbb{E}_{z\leftarrow Z}\left[\max_x \Pr[X = x|Z = z]\right]\right) = -\log\left(\mathbb{E}_{z\leftarrow Z}\left[2^{-\mathbf{H}_\infty(X|Z=z)}\right]\right)$$

where $\mathbb{E}_{z\leftarrow Z}$ denotes the expected value over $z \leftarrow Z$. We have the following lemma.

Lemma 2.1 ([DORS08])**.** *Let (X, W) be some joint distribution. Then,*

- *For any $s > 0$, $\Pr_{w\leftarrow W}[\mathbf{H}_\infty(X|W = w) \geq \mathbf{H}_\infty(X|W) - s] \geq 1 - 2^{-s}$.*
- *If Z has at most 2^ℓ possible values, then $\mathbf{H}_\infty(X|(W,Z)) \geq \mathbf{H}_\infty(X|W) - \ell$.*

The *statistical distance* between two random variables W and Z distributed over some set S is

$$\Delta(W,Z) \stackrel{def}{=} \max_{T\subseteq S}(|W(T) - Z(T)|) = \frac{1}{2}\sum_{s\in S}|W(s) - Z(s)|.$$

Note that $\Delta(W,Z) = \max_D(\Pr[D(W) = 1] - \Pr[D(Z) = 1])$, where D is a probabilistic function. We say W is ε-close to Z, denoted $W \approx_\varepsilon Z$, if $\Delta(W,Z) \leq \varepsilon$. We write $\Delta(W,Z|Y)$ as shorthand for $\Delta((W,Y),(Z,Y))$.

Extractors. An extractor [NZ96] can be used to extract uniform randomness out of a weakly-random value which is only assumed to have sufficient min-entropy. Our definition follows that of [DORS08], which is defined in terms of conditional min-entropy.

Definition 2.2 (Extractors)**.** An efficient function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (average-case, strong) (k, ε)-*extractor*, if for all X, Z such that X is distributed over $\{0,1\}^n$ and $\mathbf{H}_\infty(X|Z) \geq k$, we get

$$\Delta(\ (Z, Y, \mathsf{Ext}(X;Y))\ ,\ (Z, Y, U_m)\ ) \leqslant \varepsilon$$

where $Y \equiv U_d$ denotes the coins of Ext (called the *seed*). The value $L = k - m$ is called the *entropy loss* of Ext, and the value d is called the *seed length* of Ext.

It is well known [RTS00] that the optimal entropy loss of an extractor is $2\log(1/\varepsilon) - O(1)$, which is achieved by the famous Leftover Hash Lemma [HILL99] with seed length $d = n$. To reduce the seed length to $d = O((\log(1/\varepsilon) + \log k)\log n)$, we can also use more sophisticated extractor constructions, such as those in [GUV09, DKSS09]. Altentatively, we can extract $m = (1-\delta)k$ bits using asymptotically optimal seed length $d = O(\log(1/\varepsilon) + \log n)$ [GUV09].

Message Authentication Codes. One-time message authentication codes (MACs) use a shared random key to authenticate a message in the information-theoretic setting.

Definition 2.3 (One-time MACs)**.** A function family $\{\mathsf{MAC}_R : \{0,1\}^d \to \{0,1\}^v\}$ is an ε-*secure one-time MAC* for messages of length d with tags of length v if for any $w \in \{0,1\}^d$ and any function (adversary) $A : \{0,1\}^v \to \{0,1\}^d \times \{0,1\}^v$,

$$\Pr_R[\mathsf{MAC}_R(W') = T' \wedge W' \neq w \mid (W', T') = A(\mathsf{MAC}_R(w))] \leq \epsilon,$$

where R is the uniform distribution over the key space $\{0,1\}^\ell$.

Theorem 2.4 ([KR09])**.** *For any message length d and tag length v, there exists an efficient family of $(\lceil \frac{d}{v} \rceil 2^{-v})$-secure* MAC*s with key length $\ell = 2v$. In particular, this* MAC *is ε-secure when $v = \log d + \log(1/\epsilon)$.*

More generally, this MAC *also enjoys the following security guarantee, even if Eve has partial information E about its key R. Let (R, E) be any joint distribution. Then, for all attackers A_1 and A_2,*

$$\Pr_{(R,E)}[\mathsf{MAC}_R(W') = T' \wedge W' \neq W \mid W = A_1(E),\ (W', T') = A_2(\mathsf{MAC}_R(W), E)] \leq \left\lceil \frac{d}{v} \right\rceil 2^{v - \mathbf{H}_\infty(R|E)}.$$

(In the special case when $R \equiv U_{2v}$ and independent of E, we get the original bound.)

XOR-universal hash functions. We recall the definition of XOR-universal-hashing [CW79].

Definition 2.5 (ρ-XOR-Universal Hashing)**.** A family $\mathcal{H}$ of (deterministic) functions $h : \{0,1\}^u \to \{0,1\}^v$ is a called ρ-*XOR-universal hash family*, if for any $x_1 \neq x_2 \in \{0,1\}^u$ and any $a \in \{0,1\}^v$ we have $\Pr_{h \leftarrow \mathcal{H}}[h(x_1) \oplus h(x_2) = a] \leq \rho$. When $\rho = 1/2^v$, we say that $\mathcal{H}$ is (perfectly) *XOR-universal*. The value $\log|\mathcal{H}|$ is called the seed length of $\mathcal{H}$.

A simple construction of XOR universal hash family (for $v \leqslant u$) with seed length u sets $h_y(x) = [x \cdot y]_v$, where x and y are interpreted as elements of finite field $GF[2^u]$, $x \cdot y$ is field multiplication, and $[b]_v$ denotes the v least significant bits of b. Using standard polynomial hash [Sti94], one can also get a $\frac{u}{v \cdot 2^v}$-XOR-universal family with seed length at least v.

2.1 Privacy Amplification

We define a privacy amplification protocol (P_A, P_B), executed by two parties Alice and Bob sharing a secret $X \in \{0,1\}^n$, in the presence of an active, computationally unbounded adversary Eve, who might have some partial information E about X satisfying $\mathbf{H}_\infty(X|E) \geqslant k$. Informally, this means that whenever a party (Alice or Bob) does not reject, the key R output by this party is random and statistically independent of Eve's view. Moreover, if both parties do not reject, they must output the same keys $R_A = R_B$ with overwhelming probability.

More formally, we assume that Eve is in full control of the communication channel between Alice and Bob, and can arbitrarily insert, delete, reorder or modify messages sent by Alice and Bob to each other. In particular, Eve's strategy P_E actually defines two correlated executions (P_A, P_E) and (P_E, P_B) between Alice and Eve, and Eve and Bob, called "left execution" and "right execution", respectively. We stress that the message scheduling for both of these executions is completely under Eve's control, and Eve might attempt to execute a run with one party for several rounds before resuming the execution with another party. However, Alice and Bob are assumed to have fresh, private and independent random tapes Y and W, respectively, which are not known to Eve (who, by virtue of being unbounded, can be assumed deterministic). At the end of the left execution $(P_A(X,Y), P_E(E))$, Alice outputs a key $R_A \in \{0,1\}^m \cup \{\perp\}$, where $\perp$ is a special symbol indicating rejection. Similarly, Bob outputs a key $R_B \in \{0,1\}^m \cup \{\perp\}$ at the end of the right execution $(P_E(E), P_B(X,W))$. We let E' denote the final view of Eve, which includes E and the communication transcripts of both executions $(P_A(X,Y), P_E(E))$ and $(P_E(E), P_B(X,W)$. We can now define the security of (P_A, P_B). Our definition is based on [DLWZ11].

Definition 2.6. An interactive protocol (P_A, P_B), executed by Alice and Bob on a communication channel fully controlled by an active adversary Eve, is a (k, m, ϵ)-*privacy amplification protocol* if it satisfies the following properties whenever $\mathbf{H}_\infty(X|E) \geq k$:

1. <u>Correctness.</u> If Eve is passive, then $\Pr[R_A = R_B \wedge\ R_A \neq \perp \wedge R_B \neq \perp] = 1$.

2. <u>Robustness.</u> We start by defining the notion of *pre-application* robustness, which states that even if Eve is active, $\Pr[R_A \neq R_B \wedge\ R_A \neq \perp \wedge R_B \neq \perp] \leqslant \epsilon$.

   The stronger notion of *post-application* robustness is defined similarly, except Eve is additionally given the key R_A the moment she completed the left execution (P_A, P_E), and the key R_B the moment she completed the right execution (P_E, P_B). For example, if Eve completed the left execution before the right execution, she may try to use R_A to force Bob to output a different key $R_B \notin \{R_A, \perp\}$, and vice versa.

3. <u>Extraction.</u> Given a string $r \in \{0,1\}^m \cup \{\perp\}$, let $\mathsf{purify}(r)$ be $\perp$ if $r = \perp$, and otherwise replace $r \neq \perp$ by a fresh m-bit random string U_m: $\mathsf{purify}(r) \leftarrow U_m$. Letting E' denote Eve's view of the protocol, we require that

$$\Delta((R_A, E'), (\mathsf{purify}(R_A), E')) \leq \epsilon \quad \text{and} \quad \Delta((R_B, E'), (\mathsf{purify}(R_B), E')) \leq \epsilon$$

   Namely, whenever a party does not reject, its key looks like a fresh random string to Eve.

The quantity $k - m$ is called the *entropy loss* and the quantity $\log(1/\epsilon)$ is called the *security parameter* of the protocol.

Source Privacy. Following Bouman and Fehr [BF11], we now add the source privacy requirement for privacy amplification protocols. Our definition is actually stronger than the definition on [BF11], who only required that the final transcript E' does not reveal any information about the source X. We additionally require that the entire tuple (E', R_A, R_B) does not leak any information about the source X. Indeed, Alice and Bob might end up using their keys in application that leaks (portions of) these keys to Eve. We require that even in this case the privacy of our source X is not compromised. To define this property, we let $\mathsf{FullOutput}(X, E)$ denote the tuple (E', R_A, R_B), where Alice and Bob share a secret X and output keys R_A and R_B, respectively, and Eve starts with initial side information E and ends with final view E' at the end of the protocol.

Definition 2.7 (Source Privacy)**.** An interactive protocol (P_A, P_B), executed by Alice and Bob on a communication channel fully controlled by an active adversary Eve, is a (k, ε)-*private*, if for any two distributions (X_0, E) and (X_1, E), where $\mathbf{H}_\infty(X_0|E) \geq k$ and $\mathbf{H}_\infty(X_1|E) \geqslant k$, we have

$$\Delta(\mathsf{FullOutput}(X_0, E), \mathsf{FullOutput}(X_1, E)) \leq \varepsilon$$

Using the equivalence between entropic-security and indistinguishability [DS05], the above definition also implies that $\mathsf{FullOutput}(X, E)$ does not reveal any a priori specified function of our source X any better than what can be predicted from the initial side information E alone, provided $\mathbf{H}_\infty(X|E) \geqslant k + 2$.

We will use the following fact throughout the paper.

Fact 2.8. *Extraction with a fresh random seed is source-private.*

More precisely, consider the following simple protocol: Alice chooses a fresh random seed Y for a (k, ε) extractor Ext, sends Y to Bob (who outputs nothing), and outputs $\mathsf{Ext}(X; Y)$. This protocol is $(k, 2\varepsilon)$-private by triangle inequality, because both $\mathsf{FullOutput}(X_0, E)$ and $\mathsf{FullOutput}(X_1, E)$ are ε-close to (E, Y, U). Note, however, that if Bob also extracts from X using the received seed Y', then the protocol is no longer source-private, because Eve can give Bob a nonrandom Y' of her choice.

We remark here that for some applications, one might be interested in an interactive (n, k, m, ε) message authentication protocol iMAC as defined in [DW09] as follows:

Definition 2.9. Alice starts with a message $\mu_A \in \{0,1\}^m$ and at the end of the protocol, Bob outputs a received message $\mu_B \in \{0,1\}^m \cup \{\perp\}$. The two properties required are:

Correctness: If the adversary is passive, then for any source message μ_A, $\Pr(\mu_A = \mu_B) = 1$.

Security: If $\mathbf{H}_\infty(X|E) \geqslant k$ then, for any source message μ_A, and any active adversarial strategy of Eve, $\Pr[\mu_B \notin \{\mu_A, \perp\}] \leqslant \varepsilon$.

We note that almost all protocols that appear in the literature, starting from [RW03], and in particular, all protocols in this paper achieve interactive message authentication (as an intermediate goal before extracting R_A, R_B) with essentially the same (k, ε) for which we get pre-application robustness.

3 New Private Protocols

3.1 One Round Private Protocol

Dodis et al [DKK+12] gave a construction of robust extractors with pre-application and post-application robustness using which they gave one-round (k, m, ε)-secure privacy amplification protocols for $k > n/2 + O(\log(1/\varepsilon))$.

They give a protocol that achieves post-application robustness with entropy loss $k - m = \frac{n}{2} + O(\log(1/\varepsilon))$. In the same paper, they give another protocol that achieves pre-application robustness, but with smaller entropy loss $n-k+O(\log(1/\varepsilon))$. We observe that both their protocols are private. We argue here the source privacy of only the first protocol. The protocol is depicted as follows, where X', X'' are interpreted as elements of $\mathbb{F}_{2^{n/2}}$ and the strings $YX'+X''$ and $Y'X'+X''$ are interpreted as bitstrings in $\{0,1\}^{\frac{n}{2}}$. For any string w, by $[w]_i^j$, we denote the substring from i-th to j-th position.

Alice: X	Eve: E	Bob: X
$X = X' \| X''$		$X = X' \| X''$
Sample random $Y \in \mathbb{F}_{2^{n/2}}$		
$T = [YX' + X'']_1^v$		
	$Y, T \longrightarrow Y', T'$	
		If $T' \neq [Y'X' + X'']_1^v$ *reject*
$R_A = [YX' + X'']_{v+1}^{n/2}$		$R_B = [Y'X' + X'']_{v+1}^{n/2}$

Protocol 1: 1-round Privacy Amplification Protocol for $\mathbf{H}_\infty(X|E) > n/2$ from [DKK+12].

The privacy of this protocol follows from the following observations. It was also shown in [DKK+12] that if $\mathbf{H}_\infty(X) > n/2 + 2\log(1/\varepsilon)$, then $(Y, YX' + X'')$ is $(\varepsilon + 2^{-n/2})$-close to uniform or, equivalently, (R_A, Y, T) is $(\varepsilon + 2^{-n/2})$-close to uniform. For proving robustness of the protocol, it was shown in [DKK+12] that with probability at least $1-\varepsilon$, $R_B = R_A$ if $(Y', T') = (Y, T)$ and $R_B = \perp$, otherwise. Thus, the knowledge of R_B doesn't provide any additional information than what can be concluded from (R_A, Y, T), except with probability ε. Therefore, for any two sources X_0 and X_1 with min-entropy $k > n/2$,

$$\Delta(\mathsf{FullOutput}(X_0, E), \mathsf{FullOutput}(X_1, E)) \leq 2\varepsilon + 2^{-n/2} .$$

For the other protocol in [DKK+12] that achieves better entropy loss for pre-application robustness, the argument for source-privacy is similar. We thus get the following result.

Theorem 3.1. *For $k > n/2$, there is an explicit polynomial-time, one-round $(k, 2\varepsilon + 2^{-n/2})$-private, (k, m, ε)-secure privacy amplification protocol with pre-application robustness and entropy loss $k - m = n - k + O(\log(1/\varepsilon))$. We get post-application robustness at the cost of increasing the entropy loss to $n/2 + O(\log(1/\varepsilon))$.*

3.2 Two Round Private Protocol with Optimal Entropy Loss

In this section, we give a two round protocol that achieves optimal entropy loss $O(\log(1/\varepsilon))$ for pre-application robustness. For post-application robustness, the entropy loss is about $n/2$, but we show how to improve it to $O(\log(1/\varepsilon))$ in Section 4 at the cost of 1 additional round.

3.2.1 Non-private Protocol from [DW09]

Dodis and Wichs [DW09] showed a two-round protocol for privacy amplification with optimal (up to constant factors) entropy loss, assuming a non-malleable extractor (defined below), a regular extractor Ext (see Definition 2.2) with optimal entropy loss and any asymptotically good one-time message-authentication code MAC (see Definition 2.3).

Definition 3.2 (Non-Malleable Extractors)**.** An efficient function $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (average-case) (k, ε)-*non-malleable extractor*, if for all $X \in \{0,1\}^n, Z \in \mathcal{Z}$ such that $\mathbf{H}_\infty(X|Z) \geq k$, and any function $\mathcal{A} : \{0,1\}^d \times \mathcal{Z}$ such that $\mathcal{A}(y,z) \neq y$ for all y, z, we have

$$\Delta(\mathsf{nmExt}(X;Y), U_m \mid Z, Y, \mathsf{nmExt}(X;\mathcal{A}(Y,Z)) \leqslant \varepsilon$$

where $Y \equiv U_d$ denotes the coins of nmExt.

Using these, their protocol is depicted as Protocol 2.

Alice: X	Eve: E	Bob: X
Sample random Y		
	$Y \longrightarrow Y'$	
		Sample random W' $K' = \mathsf{nmExt}(X;Y')$ $T' = \mathsf{MAC}_{K'}(W')$ Set final $R_B = \mathsf{Ext}(X;W')$
	$(W,T) \longleftarrow (W',T')$	
$K = \mathsf{nmExt}(X;Y)$ **If** $T \neq \mathsf{MAC}_K(W)$ *reject* Set final $R_A = \mathsf{Ext}(X;W)$		

Protocol 2: 2-round Privacy Amplification Protocol for $\mathbf{H}_\infty(X|E) > n/2$ from [DW09].

There have been some recent constructions of non-malleable extractors [DLWZ11, Li12c, CRS12]. Using the non-malleable extractor from [Li12c] with seed length $d = n$ and $m = \Omega(n)$ for $k \geqslant \frac{n}{2}(1-\alpha) + O(\log(1/\varepsilon))$ for some small universal constant α, we get the following:

Theorem 3.3 ([DW09, Li12c])**.** *Assuming error $\varepsilon < 1/n$ and min-entropy $k \geq \frac{n}{2}(1-\alpha) + \Theta(\log(1/\varepsilon))$, there exists is a polynomial-time, two-round (k, m, ε)-secure privacy amplification protocol with entropy loss $k - m = O(\log(1/\varepsilon))$ for pre-application robustness.*

Observe that this protocol is not source-private: if Eve uses $Y' \neq Y$, then K', and therefore T', may contain useful information about X.

3.2.2 Our Two Round Private Protocol

Idea: Our protocol, depicted as Protocol 3 makes the protocol of [DW09] (given in Section 3.2.1) private, using the same idea as [BF11]: we will apply a one-time pad P' to the tag sent by Bob in the second round, T', where the pad P' is derived from X. However, *how* the pad is derived will be different from [BF11]. Specifically, Bob will derive the pad using a fresh random seed S' to extract it from X; he will then send S' to Alice. Source privacy is now clear from Fact 2.8 (assuming robustness holds, which ensures that $R_A = R_B$ or $\perp$, and Eve knows with high probability whether $R_A = \perp$): Eve sees only random seeds Y, W', and S', the value R_B that was extracted from X using a random W', and the value C' that was extracted from X using S' and then shifted by some T' (of which S' is independent). However, robustness of privacy amplification itself is not obvious anymore. We show that privacy amplification can still be achieved as long as the extractor to obtain the one-time pad is what we call an *adaptive non-malleable* extractor, which we define below.

Alice: X	Eve: E	Bob: X
Sample random Y		
	$Y \longrightarrow Y'$	
		Sample random $W', S' \neq Y'$ $K' = \mathsf{anmExt}(X; Y')$ $T' = \mathsf{MAC}_{K'}(W')$ $P' = \mathsf{anmExt}_{1..\ell}(X; S')$ $C' = T' \oplus P'$ Set final $R_B = \mathsf{Ext}(X; W')$
	$(W, S, C) \longleftarrow (W', S', C')$	
If $Y = S$ *reject* $K = \mathsf{anmExt}(X; Y)$ $P = \mathsf{anmExt}_{1..\ell}(X; S)$ **If** $C \oplus P \neq \mathsf{MAC}_K(W)$ *reject* Set final $R_A = \mathsf{Ext}(X; W)$		

Protocol 3: New 2-round Source-Private Privacy Amplification Protocol for $\mathbf{H}_\infty(X|E) > n/2$

Adaptive Non-malleable Extractors: As mentioned above, we will need a stronger notion of non-malleability than used in previous works, in which $\mathcal{A}$ is allowed to see Y, Z, and additionally either $\mathsf{anmExt}(X; Y)$ or $R \equiv U_m$ before producing the modified seed Y'.

Definition 3.4. An efficient function $\mathsf{anmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (average-case) (k, ε)-*adaptive non-malleable extractor*, if for all $X \in \{0,1\}^n, Z \in \mathcal{Z}$ such that $\mathbf{H}_\infty(X|Z) \geq k$, and any function $\mathcal{A} : \{0,1\}^d \times \mathcal{Z} \times \{0,1\}^m \to \{0,1\}^d$ such that $\mathcal{A}(y, z, r) \neq y$ for all (y, z, r), we have

$$\Delta\Big(\big(P, \mathsf{anmExt}(X; \mathcal{A}(Y, Z, P))\big), \big(R, \mathsf{anmExt}(X; \mathcal{A}(Y, Z, R))\big) \mid Y, Z\Big) \leqslant \varepsilon \ ,$$

where $Y \equiv U_d$ denotes the seed for anmExt, $R \equiv U_m$, $P = \mathsf{anmExt}(X; Y)$.

Dodis and Yu [DY13] informally introduced the notion of adaptive non-malleable extractors as a special case of a family of (q, δ)-wise independent hash functions. They constructed a non-malleable extractor for $k \geqslant \frac{n}{2} + O(\log(1/\varepsilon))$, and observed that it is also an adaptive non-malleable extractor. (The same construction was independently discovered by [Li12c], but the proof there does not immediately give adaptive non-malleability.)

Here we show how to get an adaptive non-malleable extractor from any non-malleable extractor.

Lemma 3.5. *A non-malleable (k, ε) extractor* $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is $(k, \varepsilon 2^m)$ adaptive non-malleable.*

Proof. First, observe that the definition of non-malleable extractor (Definition Definition 3.2 is equivalent to (a seemingly more stringent) definition in which $\mathcal{A}$ is a randomized algorithm that takes randomness S, and we require

$$\Delta(P, R \mid Z, Y, S, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, S)) \leqslant \varepsilon \ ,$$

where $P = \mathsf{nmExt}(X; Y)$, and $R \equiv U_m$. The equivalence follows by the usual argument of hard-wiring the "best" randomness S into $\mathcal{A}$ and the distinguisher for statistical distance. In other words, given any randomized algorithm $\mathcal{A}$ that takes randomness S as input, we can replace it with a deterministic algorithm by fixing $S = s$ that maximizes the statistical distance. We will use this definition for the purpose of this proof.

Using notation of Definition 3.4, take some $\mathcal{A}$ and D and let

$$\begin{aligned}
\Pr[D(Z, Y, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, P), P, P) = 1] &= p_1\,, \\
\Pr[D(Z, Y, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, R)), R, R) = 1] &= p_2\,,
\end{aligned}$$

Suppose $p_1 - p_2 > \varepsilon 2^m$. Let $\mathcal{A}'(Y)$ be a randomized function that chooses a uniform S and computes $\mathcal{A}(Y, S)$. Define D' as a distinguisher that checks whether the last two components of its input are equal; if they are equal, then it invokes D, otherwise it outputs 0. Then

$$\Pr[D'(Z, Y, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, P)), S, P) = 1] = p_1 2^{-m}\,,$$

and

$$\Pr[D'(Z, Y, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, R)), S, R) = 1] = p_2 2^{-m}\,.$$

This D' and A' violate (k, ε) adaptive non-malleability of nmExt. Therefore, for any $\mathcal{A}$,

$$\Delta\Big(\big(P, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, P)\big); \big(R, \mathsf{nmExt}(X; \mathcal{A}(Y, Z, R)\big) \mid Y, Z\Big) \leqslant \varepsilon 2^m\,.$$

$\square$

We use this result along with the ε'-secure non-malleable extractor from [Li12c] for $k \geqslant \frac{n}{2}(1 - \alpha) + O(\log(1/\varepsilon))$ with output length $m = \Theta(\log(1/\varepsilon))$ such that $\varepsilon' \cdot 2^m \leqslant \varepsilon$ to get the following result.

Corollary 3.6. *There exists an explicit (k, ε) adaptive non-malleable extractor for $k \geqslant \frac{n}{2}(1-\alpha) + O(\log(1/\varepsilon))$ that uses seed of length n, and has output length $\Theta(\log(1/\varepsilon))$.*

Also, we can use the result proving the existence of non-malleable extractors from [DW09] for $k = \Omega(\log(1/\varepsilon))$ to get the following result.

Corollary 3.7. *There exists an (k, ε) adaptive non-malleable extractor for $k = \Omega(\log(1/\varepsilon))$ that uses seed of length n, and has output length $\frac{k}{2} - O(\log(1/\varepsilon))$.*

The Protocol: Let $\varepsilon' = \varepsilon/7$. We will need the following building blocks:

- Let $\mathsf{anmExt} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^{2\ell}$ be a (τ, ε')-adaptive non-malleable extractor, for $\ell = O(\log(1/\varepsilon))$, $\tau \leqslant k - 3\ell$, and $t \geqslant 2\log(1/\varepsilon)$. Let $\mathsf{anmExt}_{a..b}(X; Z)$, where $1 \leqslant a \leqslant b \leqslant 2\ell$, denote the sub-string of extracted bits from bit position a to bit position b.
- Let $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a $(k - 3\ell, \varepsilon')$-extractor with $d = O(\log n \log(1/\varepsilon))$.
- Let MAC be an ε'-secure one-time MAC for d-bit messages, whose key length is 2ℓ (the output length of nmExt). Using the construction from Theorem 2.4, we set the tag length ℓ.

Using the above building blocks, the protocol is depicted as Protocol 3. We obtain the following result.

Theorem 3.8. *Let $2^{-n/4} < \varepsilon < 1/n$, and $\varepsilon' = \varepsilon/7$. Given a (τ, ε')-adaptive non-malleable extractor, for $k > \tau + \Theta(\log(1/\varepsilon))$ and output length $\Theta(\log(1/\varepsilon))$, there exists an explicit polynomial-time, two-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with pre-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$. Furthermore, we get post-application robustness at the cost of increasing the entropy loss to $\tau + O(\log(1/\varepsilon))$.*

Proof. We first argue $5\varepsilon'$ pre-application robustness for entropy loss $k - m = O(\log(1/\varepsilon))$. To have any chance of breaking robustness, Eve must give Alice $W \neq W'$ and $S \neq Y$, because if $W = W'$ or $S = Y$, then either $R_A = R_B$ or $R_A = \perp$. Thus assume $W \neq W'$ and $S \neq Y$.

If $S = S'$, then robustness follows from the robustness of Protocol 2. To see this, suppose Eve breaks robustness of Protocol 3 while maintaining $S = S'$. Then we will build Eve$'$ to break robustness of Protocol 2, except with entropy of X reduced by the length of P' (which is $O(\log(1/\varepsilon))$. Specifically, we will let the knowledge E' of Eve$'$ include the value $P' = \mathsf{nmExt}(X; S')$ on a random S' (as well as S' itself). The reduction is straightforward. For the first message Y, Eve$'$ gives up if $Y = S'$; otherwise, she will compute Y' the same way as Eve, and send it to Bob. For the second message, Eve$'$ will compute $C' = T' \oplus P'$. She will give (W', S', C') to Eve, who will return $(W, S = S', C)$; Eve$'$ will then compute $T = C \oplus P'$ and send (W, T) to Alice. It is easy to see that Eve$'$ will succeed in violating robustness of Protocol 2 whenever Eve succeeds, unless $Y' = S$, which happens with probability 2^{-t}.

If $W \neq W'$, $S \neq Y$, and $S \neq S'$, then given $E, Y, W', S', T' \oplus P'$, the adversary needs to compute $W, S, P \oplus \mathsf{MAC}_K(W)$ in order to break robustness. Consider a slightly modified adversarial game, in which Eve received P upon specifying S, and only then has to specify W and $\mathsf{MAC}_K(W)$. By a straigthforward reduction in this modified game Eve is no weaker: whatever she can accomplish without knowing P, she can also accomplish in this game. Thus, to break robustness, Eve needs to compute $W, \mathsf{MAC}_K(W)$ with probability greater than $5\varepsilon'$ given $E, Y, W', S', T' \oplus P', P$.

Note that S' is equal to Y with probability at most $1/(2^t - 1) < \varepsilon'$. This implies that conditioned on the event that S' is not equal to Y, Eve succeeds with probability at least $4\varepsilon'$.

Since Y, W' are independent of X, we have using Lemma 2.1,

$$\mathbf{H}_\infty(X|E, Y, W', T', K) \geqslant k - 3\ell \,.$$

Thus, using the adaptive non-malleability property of anmExt and the fact that W, S are functions of $E, Y, W', S', T' \oplus P'$, we have that the statistical distance between the joint distribution $(T' \oplus P', E, Y, W', S', S, P, W, \mathsf{MAC}_K(W))$ and $(U_\ell, E, Y, W', S', S, P, W, \mathsf{MAC}_K(W))$ is at most ε'. Since the adversary can herself simulate U_ℓ, this implies that there is an adversary that succeeds in computing $W, \mathsf{MAC}_K(W)$ with probability greater than $3\varepsilon'$ given E, Y, W', S', S, P.

For any fixed W', S', S is a deterministic function of Y and E. Using non-malleability property of anmExt, we have that K is ε'-close to uniform given E, Y, W', S', S, P. Thus, in order to win, the adversary must output $W, \mathsf{MAC}_K(W)$ for a random key K with probability more than $2\varepsilon'$, which leads to a contradiction.

For post-application robustness, we must analyze the case where the adversary also gets R_B in addition to the final transcript E'. Note that we can do essentially the same analysis as for pre-application robustness, except that the non-malleable extractor should be secure even given R_B, and so the entropy bound that we need is $\tau \leqslant k - 3\ell - m$, or in other words, $k - m \geqslant \tau + 3\ell$.

The extraction property follows easily from robustness. Note that Y, S' are independent of X. Thus, using Lemma 2.1,

$$\mathbf{H}_\infty(X|E, Y, S', P', K') \geqslant k - 3\ell \,.$$

Thus, using the definition of Ext, we have that R_B is ε'-close to uniform given the transcript of Eve, E'. Also, with probability $1 - 5\varepsilon'$, $R_A = R_B$ or $R_A = \perp$. Thus, conditioned on the event that

$R_A = R_B$ or $R_A = \perp$, R_A is ε' close to $\mathsf{purify}(R_A)$. Thus,

$$\Delta((R_A, E'), (\mathsf{purify}(R_A), E')) \leq 5\varepsilon' + \varepsilon' = 6\varepsilon' \ .$$

The source privacy follows easily from the following observations: R_B is ε'-close to uniform given the transcript E, Y, W', S', C'. Also, P', and hence C' is ε'-close to uniform given E, Y, W', S'. Finally note that $R_B = R_A$ if $W' = W$ and $R_A = \perp$, otherwise except with probability at most $5\varepsilon'$. Thus, the knowledge of R_B doesn't provide any additional information than what can be concluded from (R_A, E'), except with probability $5\varepsilon'$. Thus

$$\Delta(\mathsf{FullOutput}(X_0, E), \mathsf{FullOutput}(X_1, E)) \leq 5\varepsilon' + \varepsilon' + \varepsilon' = 7\varepsilon' \ .$$

□

We can instantiate the above result using the adaptive non-malleable extractor obtained by using Corollary 3.6 to get the following result.

Corollary 3.9. *There exists a universal constant $\alpha > 0$, such that for $k > n/2(1-\alpha)$, there exists an explicit polynomial-time, two-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with pre-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$. We get post-application robustness at the cost of increasing the entropy loss to $n/2(1-\alpha) + O(\log(1/\varepsilon))$.*

Similarly, using Corollary 3.7, we get the following result.

Corollary 3.10. *For $k = \Omega(\log(1/\varepsilon))$, there exists a two-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with post-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$.*

3.3 Privacy using Extractors and XOR-Universal Hashing

In this section, we use a ρ-XOR universal hash function family to construct a 4-round protocol for private privacy amplification, given any 2 round privacy amplification protocol of the form Protocol 4, where the string sent in the first round is sampled independent of X. We note that all known 2 round protocols in the literature are of this generic form.

Alice: X	Eve: E	Bob: X
Sample random Y		
	$Y \longrightarrow Y'$	
		Sample random W' $K' = f_1(X, Y')$ $T' = f_2(K', W')$ Set final $R_B = g(X, W')$
	$(W, T) \longleftarrow (W', T')$	
$K = f_1(X, Y)$ **If** $T \neq f_2(K, W)$ *reject* Set final $R_A = g(X, W)$		

Protocol 4: A Generic 2-round Privacy Amplification Protocol

Let $\ell = \log(1/\varepsilon)$. Let $\mathcal{H}$ be a ε-XOR universal family of hash functions from $\{0,1\}^{|T|}$ to $\{0,1\}^{2\ell}$, and let $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^{2\ell}$ be a $(k - 2\ell - 2|K| - |R_B|, \varepsilon)$ extractor. Using these, our protocol is depicted as Protocol 5.

For proving the security of our generic transformation, we need the following result.

Alice: X	Eve: E	Bob: X
Sample random Y		
	$Y \longrightarrow Y'$	
		Sample random W', S'
	$W, S \longleftarrow W', S'$	
Sample random h		
	$h \longrightarrow h'$	
		$K' = f_1(X, Y')$ $T' = f_2(K', W')$ $C' = h'(T') \oplus \mathsf{Ext}(X; S')$ Set final $R_B = g(X, W')$
	$C \longleftarrow C'$	
$K = f_1(X, Y)$ $T = f_2(K, W)$ **If** $C \neq h(T) \oplus \mathsf{Ext}(X; S)$ *reject* Set final $R_A = g(X, W)$		

Protocol 5: A Generic 4-round Private Privacy Amplification Protocol

Lemma 3.11. *Given any random variables $A \in \{0,1\}^u, B \in \{0,1\}^v, C \in \mathcal{C}$, let $H : \{0,1\}^u \to \{0,1\}^v$ be a function chosen uniformly at random from a family of ρ-XOR-universal hash functions $\mathcal{H}$. Then,*

$$\mathbf{H}_\infty(H(A) \oplus B | H, C) \geqslant -\frac{1}{2}\log(\rho + 2^{-\mathbf{H}_\infty(A|C)}) .$$

Proof. We define the probability of guessing a random variable X conditioned on another random variable Z as $\mathsf{Pred}(X|Z) = 2^{-\mathbf{H}_\infty(X|Z)}$, and collision probability of X conditioned on Z as

$$\mathsf{Col}(X|Z) = \mathbb{E}_{z \leftarrow Z}\left[\sum_x \Pr[X = x | Z = z]^2\right] = \mathbb{E}_{z \leftarrow Z}\left[\Pr[X = X' | Z = z]\right] ,$$

where for any $z \in \mathrm{Support}(Z)$, $X'|_{Z=z}$ is independent and identically distributed as $X|_{Z=z}$. It is easy to see that

$$\mathsf{Col}(X|Z) \leqslant \mathsf{Pred}(X|Z) \leqslant \sqrt{\mathsf{Col}(X|Z)} . \tag{1}$$

Using (both inequalities in) Equation 1, we see that it is enough to prove that

$$\mathsf{Col}(H(A) \oplus B | H, C) \leqslant \rho + \mathsf{Col}(A|C) .$$

For any h, c, let $(A', B')|_{H=h,C=c}$ be independent and identically distributed as $(A, B)|_{H=h,C=c}$. We have

$$\begin{aligned}
\mathsf{Col}(H(A) \oplus B | H, C) &= \mathbb{E}_{c \leftarrow C, h \leftarrow H}\left[\Pr(H(A) \oplus B = H(A') \oplus B' | C = c)\right] \\
&= \mathbb{E}_{c \leftarrow C}\left[\Pr(H(A) \oplus B = H(A') \oplus B' | C = c)\right] \\
&\leqslant \mathbb{E}_{c \leftarrow C}\left[\Pr(A = A' | C = c)\right] \ + \\
&\quad\ \mathbb{E}_{c \leftarrow C}\left[\Pr(H(A) \oplus H(A') = B \oplus B' \wedge A \neq A' | C = c)\right] \\
&\leqslant \mathsf{Col}(A|C) + \rho .
\end{aligned}$$

$\square$

Theorem 3.12. *Let Protocol 4 be a 2-round $(k-u, m, \varepsilon)$-secure privacy amplification protocol with pre- (resp. post-) application robustness for $k - |T| - 2|K| - |R_B| \geqslant 2\ell$. Then Protocol 5 is a 4-round $(k, m, O(\sqrt{\varepsilon}))$-secure $(k, O(\sqrt{\varepsilon}))$-private privacy amplification protocol with pre- (resp. post-) application robustness.*

Proof. The correctness of Protocol 5 follows trivially from the correctness of Protocol 4.

We argue post-application robustness of Protocol 5 assuming post-application robustness of Protocol 4. The argument for pre-application robustness is similar, except that R_B is not revealed to the adversary. For post-application robustness, we must analyze the case where the adversary also gets R_B in addition to the final transcript E'. To have any chance of breaking robustness, Eve must choose $W \neq W'$ and C such that $R_A \notin \{R_B, \perp\}$, and so we assume that this is the case. Eve succeeds if she can compute $h(T) \oplus \mathsf{Ext}(X;S)$, given $R_B, Y, S', h, W', h'(T') \oplus \mathsf{Ext}(X;S')$. Let the success probability of Eve be ε^*. Observe that since Eve fully controls the channel, she can interact with Alice and Bob separately and does not have to respect the message order specified by the protocol. Alice and Bob, however, do respect the message order specified by the protocol. We consider two cases.

CASE 1: Eve sends Y' to Bob after receiving h from Alice. In this case, S' is independent of h, S, K, K', W. Also, since X is independent of Y, S', h, W', we have that

$$\begin{aligned} \mathbf{H}_\infty(X|R_B, Y, h, W', h'(T'), h(T), \mathsf{Ext}(X,S)) &\geqslant \mathbf{H}_\infty(X|R_B, K', K, \mathsf{Ext}(X;S)) \\ &\geqslant k - 2\ell - 2|K| - |R_B| \,. \end{aligned}$$

Thus, using the fact that Ext is a strong extractor, we have that $h'(T') \oplus \mathsf{Ext}(X;S')$ is ε-close to uniform given $R_B, Y, S', h, W', h(T), \mathsf{Ext}(X;S)$. So, Eve must be able to compute $h(T) \oplus \mathsf{Ext}(X;S)$ given $R_B, Y, S', h, W', U_{2\ell}$ with probability $\varepsilon^* - \varepsilon$.

Therefore, the success probability of the adversary in computing $h(T) \oplus \mathsf{Ext}(X;S)$ given R_B, Y, h, W' is at least $\varepsilon^* - \varepsilon$, since the adversary can simulate S' and $U_{2\ell}$ herself. Using Lemma 3.11 and the security of Protocol 4, we have that $\varepsilon^* - \varepsilon = O(\sqrt{\varepsilon})$), which implies $\varepsilon^* = O(\sqrt{\varepsilon})$).

CASE 2: Eve sends Y' to Bob before receiving h from Alice. In this case, h is independent of S, W, Y'. We give the adversary additional power by assuming that the adversary gets T' and $\mathsf{Ext}(X;S')$ for free. Thus, Eve succeeds in computing $h(T) \oplus \mathsf{Ext}(X;S')$, given R_B, Y, S', h, W', T', $\mathsf{Ext}(X;S')$ with probability ε^*. We have that $\mathbf{H}_\infty(X|\mathsf{Ext}(X;S')) \geqslant k - 2\ell$. Using the security of Protocol 4, we have that

$$\mathbf{H}_\infty(T|Y, W', R_B, \mathsf{Ext}(X,S'), S', T') \geqslant \log 1/\varepsilon \,.$$

Thus, using Lemma 3.11, we get that $\varepsilon^* = O(\sqrt{\varepsilon})$).

The extraction property follows easily from the extraction property of protocol 4 and the fact that $\mathbf{H}_\infty(X|\mathsf{Ext}(X;S') \geqslant k - 2\ell$. As usual, we give Eve additional power by assuming that she gets $\mathsf{Ext}(X;S')$ and T'. Thus, using the extraction property of protocol 4, R_B is ε-close to uniform given Eve's view, since in addition to Protocol 4, Eve sees h, S' (which are independent of R_B) and $\mathsf{Ext}(X;S')$ which is of length at most 2ℓ. Also, with probability $1 - O(\sqrt{\varepsilon})$, $R_A = R_B$ or $R_A = \perp$. Thus, conditioned on the event that $R_A = R_B$ or $R_A = \perp$, R_A is ε-close to $\mathsf{purify}(R_A)$. Therefore,

$$\Delta((R_A, E'), (\mathsf{purify}(R_A), E')) \leq \varepsilon + O(\sqrt{\varepsilon}) = O(\sqrt{\varepsilon}) \,.$$

The source privacy follows easily from the following observations: R_B is ε-close to uniform given Eve's view, and Y, S', W', h are independent of source X. Also, $h'(T') \oplus \mathsf{Ext}(X; S')$ is ε-close to uniform given R_B, Y, S', h, W', as argued in CASE 1, above. Furthermore, as shown above, $R_B = R_A$ if $W' = W$ and $R_A = \perp$, otherwise except with probability at most $O(\sqrt{\varepsilon})$. Thus, the knowledge of R_B doesn't provide any additional information than what can be concluded from (R_A, E'), except with probability $O(\sqrt{\varepsilon})$. Therefore,

$$\Delta(\mathsf{FullOutput}(X_0, E), \mathsf{FullOutput}(X_1, E)) \leq O(\sqrt{\varepsilon}) + \varepsilon = O(\sqrt{\varepsilon}) .$$

□

We apply this generic transformation to Li's recent 2 round (k, ε)-secure privacy amplification protocol for $k = \Omega(\log^2(1/\varepsilon))$, that achieves entropy loss $O(\log(1/\varepsilon))$ for pre-application robustness, and $O(\log^2(1/\varepsilon))$ for post-application robustness [Li12b]. We get the following result.

Corollary 3.13. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time, 4-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with pre-application robustness and entropy loss $L = k - m = O(\log(1/\varepsilon))$. We get post-application robustness with entropy loss $O(\log^2(1/\varepsilon))$.*

In Section 4, we will see how to get a 5-round private privacy amplification protocol with post-application robustness and entropy loss $O(\log(1/\varepsilon))$.

Remark 1: We can apply the generic transformation to the 2 round construction of [DW09] for $k \geqslant \log^2(1/\varepsilon)$ to decrease the residual entropy loss from $O(\log^2(1/\varepsilon))$ to $O(\log(1/\varepsilon))$, and simultaneously achieving source privacy. In Section 5, we will give a 2 round to 2 round generic transformation that decreases the residual entropy loss to $O(\log(1/\varepsilon))$ but does not achieve source privacy.

Remark 2: We note that our result in this section also achieves 4 round private "liveness test" with optimal residual entropy loss. Liveness tests (aka "identification schemes") are similar to iMAC for message space of cardinality 1 (except that they must be interactive to prevent replay attacks). The standard protocol is to send an extractor seed Y and respond with $\mathsf{Ext}(X; Y)$ (or, send, a random hash function chosen from an almost universal hash function family and respond with $h(X)$). But none of these schemes achieves source privacy. Using our transformation we can achieve source privacy for liveness tests. Though we need 4 rounds instead of 2, we can still have residual entropy loss $O(\log(1/\varepsilon))$, optimal upto constant factors.

4 From Pre-application to Post-application Robustness

In this section, we show a generic transformation from a t-round privacy amplification protocol $\mathcal{P}$ that achieves pre-application robustness to a $(t+1)$-round protocol $\mathcal{P}'$ that achieves post-application robustness. The transformation can be described as follows.

Let $\ell = \log(1/\varepsilon)$. Without loss of generality, assume that the last message in $\mathcal{P}$ was sent from Bob to Alice. Let $\widetilde{R}_A, \widetilde{R}_B$ denote the first u bits of the keys computed by Alice and Bob, respectively (Set $\widetilde{R}_A = \perp$ if Alice rejects, and $\widetilde{R}_B = \perp$ if Bob rejects). We need a $(k - O(\ell), \varepsilon)$-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ and an ε-secure one-time MAC for d-bit messages, whose key length is u. Using these, the $(t+1)$-round protocol is depicted as Protocol 6.

Alice: X	Eve: E	Bob: X
	$\longleftarrow$	
	$\ldots$	
	Protocol $\mathcal{P}$	
	$\longrightarrow$	
	$\ldots$	
	$\longleftarrow$	
Sample random S $T = \mathsf{MAC}_{\widetilde{R}_A}(S)$ Set $R_A = \mathsf{Ext}(X; S)$		
	$S, T \longrightarrow S', T'$	
		If $T' \neq \mathsf{MAC}_{\widetilde{R}_B}(S')$ *reject* Set final $R_B = \mathsf{Ext}(X; S')$

Protocol 6: $(t+1)$-round Privacy Amplification Protocol $\mathcal{P}'$ with post-application robustness.

Theorem 4.1. *If Protocol $\mathcal{P}$ is (k, m, ε)-secure privacy amplification protocol with pre-application robustness and residual entropy $k - O(\log(1/\varepsilon))$, then Protocol $\mathcal{P}'$ is a $(k, m - O(\log(1/\varepsilon)), O(\varepsilon))$ secure privacy amplification protocol with post-application robustness. Additionally, if $\mathcal{P}$ is (k, ε) private, then $\mathcal{P}'$ is $(k, O(\varepsilon))$ private.*

Proof. Let $|R_A| = |R_B| = r$ (if Alice and Bob do not reject).

The correctness follows trivially from the correctness of Protocol $\mathcal{P}$.

We show 4ε post-application robustness of $\mathcal{P}'$. We assume that either one of $\widetilde{R}_A$, $\widetilde{R}_B$ is $\perp$, or $\widetilde{R}_A = \widetilde{R}_B$. By (pre-application) robustness of $\mathcal{P}$, this happens with probability at least $1-\varepsilon$. If one of $\widetilde{R}_A$ or $\widetilde{R}_B$ is $\perp$, then, either R_A or R_B is $\perp$. So, we assume that $\widetilde{R}_A = \widetilde{R}_B = \widetilde{R} \neq \perp$. Thus, to have any chance of breaking post-application robustness, the adversary must set $S' \neq S$, and needs to compute $\mathsf{MAC}_{\widetilde{R}}(S')$ given Eve's view E' of protocol $\mathcal{P}$, S, $\mathsf{MAC}_{\widetilde{R}}(S)$, and R_A with probability 3ε. By using that Ext is a strong extractor and the fact that X has entropy $k - O(\log(1/\varepsilon))$ given $E', \widetilde{R}$, we have that R_A is ε-close to U_r given $E', S, \mathsf{MAC}_{\widetilde{R}}(S), \mathsf{MAC}_{\widetilde{R}}(S')$. This implies that, since Eve can simulate U_r herself, she should be able to compute $\mathsf{MAC}_{\widetilde{R}}(S')$ with probability 2ε given $E', S, \mathsf{MAC}_{\widetilde{R}}(S)$. By the security of MAC, this is impossible.

Extraction: If $R_A \neq \perp$, then as seen above, R_A is ε-close to uniform given $E', S, \mathsf{MAC}_R(S)$. This implies the extraction property for Alice.

The extraction property for Bob requires a little more work. Let $\widetilde{R} = \widetilde{R}_B$. We know from the extraction property of $\mathcal{P}$ that

$$\Delta((R, E'), (\mathsf{purify}(R), E')) = \Delta((R, E'), (U_u, E')|R \neq \perp) \cdot \Pr(R \neq \perp) \leqslant \varepsilon \ . \tag{2}$$

Let $\Delta((R, E'), (U_u, E')|R \neq \perp) = \beta$. By robustness, we have that with probability $1 - O(\varepsilon)$ either one of R_A, R_B is $\perp$, or $R_A = R_B$. We assume that this is the case. Further, we assume that $R_B \neq \perp$. This implies that $R \neq \perp$. Thus, R is β-close to uniform given E'. If $R_A \neq \perp$, then $R_A = R_B$, and R_A, and hence R_B is ε-close to uniform. If $R_A = \perp$, then nothing is sent in the $(t+1)$-th round by Alice, and so Eve must guess $\mathsf{MAC}_R(S')$ correctly which happens with probability at most $\beta + \varepsilon$

by the security of MAC, and the fact that R is β-close to uniform. Therefore,

$$\begin{aligned}
\Delta((R_B, E'), (\mathsf{purify}(R_B), E')) &= \Delta((R_B, E'), (U_r, E')|R_B \neq \perp) \cdot \Pr(R_B \neq \perp) \\
&\leqslant \varepsilon + \Delta((R_B, E'), (U_r, E')|R_B = R_A \neq \perp) \cdot \Pr(R_B = R_A \neq \perp) + \\
&\quad \Delta((R_B, E'), (U_r, E')|R_A = \perp, R_B \neq \perp) \cdot \Pr(R_A = \perp, R_B \neq \perp) \\
&\leqslant \varepsilon + \varepsilon \cdot \Pr(R_B = R_A \neq \perp) + (\beta + \varepsilon) \cdot \Pr(R_A = \perp, R_B \neq \perp) \\
&\leqslant 2\varepsilon + \beta \cdot \Pr(R \neq \perp) + \varepsilon \leq 4\varepsilon\ ,
\end{aligned}$$

where we used the fact that $\Pr(R_A = \perp, R_B \neq \perp) \leqslant \Pr(R \neq \perp)$, and equation 2.

Source Privacy: We need to show that $\mathsf{FullOutput}(X, E) = (R_A, R_B, E', S, \mathsf{MAC}_{\widetilde{R}_A}(S))$ is $O(\varepsilon)$-close to $\mathsf{FullOutput}(Y, E)$, from any other source Y such that $\mathbf{H}_\infty(Y|E) \geqslant k$. In fact a stronger statemnet holds, namely, $\mathsf{FullOutput}^\star(X, E) := (R_A, R_B, E', S, \widetilde{R}_A, \widetilde{R}_B)$ is $O(\varepsilon)$-close to $\mathsf{FullOutput}^\star(Y, E)$. By post-application robustness, we know that with probability $1 - \varepsilon$ either one of R_A, R_B is $\perp$, or $R_A = R_B$. We assume that this is the case. If $R_A \neq \perp$, then we know that R_A is ε-close to uniform given $E', S, \widetilde{R}_A, \widetilde{R}_B$. Also, given E', S, $\widetilde{R}_A, \widetilde{R}_B$, it is easy to check whether any of R_A or R_B is $\perp$. Thus, the knowledge of R_A, R_B does not provide any additional information, except with probability at most ε. Then, the source privacy follows from the source privacy of $\mathcal{P}$. □

Applying this generic transformation on the two round and four round protocols given by Corollary 3.9, 3.13 give us the following:

Corollary 4.2. *There exists a universal constant $\alpha > 0$, such that for $k > n/2(1-\alpha)$, there exists an explicit polynomial-time, 3-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with post-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$.*

Corollary 4.3. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time, 5-round (k, ε)-private, (k, m, ε)-secure privacy amplification protocol with post-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$.*

Also, we can apply this generic transformation to the (non-private) two round protocol of [Li12b] that achieves pre-application robustness for $k = O(\log^2(1/\varepsilon))$ with entropy loss $O(\log(1/\varepsilon))$ to get the following result.

Corollary 4.4. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time, 3-round (k, m, ε)-secure privacy amplification protocol with post-application robustness and entropy loss $k - m = O(\log(1/\varepsilon))$.*

5 Increasing Residual Entropy

In this section we consider the task of preserving as much entropy as possible in the weak source X after the execution of a privacy amplification protocol. This is an arguably natural goal, and in particular it has implications in the Bounded Retrieval Model where there is a huge weak source which one wants to use in many sequential protocol executions (see section 6). Formally, we define the *residual entropy* of an interactive protocol using a weak source X as $\min_{E'}(\mathbf{H}_\infty(X \mid E'))$ where E' is the adversary's view after the protocol (i.e. E' contains the initial side-information E and the protocol transcript). We refer to $\mathbf{H}_\infty(X \mid E) - \min_{E'}(\mathbf{H}_\infty(X \mid E'))$ as the loss in residual entropy.

The main result of this section is the following theorem, which transforms a given privacy amplification protocol with post-application robustness into one that achieves loss in residual entropy $O(\log(1/\varepsilon)$, i.e. linear in the security parameter, which is optimal up to constant factors.

Theorem 5.1. *Assume that there is a 2-round (k, m, ε)-secure privacy amplification protocol with post-application robustness in which the first message is independent of the (n, k)-source X and we have $\log n = O(\log(1/\varepsilon))$, $\varepsilon \geq 2^{-m/C}$, and $k \geq C \log(1/\varepsilon)$ for sufficiently large C.*

Then there is a 2-round (k', m', ε')-secure privacy amplification protocol with residual entropy $\geq k' - O(\log(1/\varepsilon'))$ provided that $k' \geq k + C' \log(1/\varepsilon)$ and $\varepsilon' \geq \varepsilon^{1/C'}$ for sufficiently large C', and $m' = k' - O(\log(1/\varepsilon'))$ for pre-application robustness or $m' = k' - k - O(\log(1/\varepsilon'))$ for post-application robustness.

For the remainder of this section, every protocol under consideration will have the property that the first message sent (which is always sent by Alice) is independent of the weak source X; we avoid restating this in each theorem for the sake of brevity.

5.1 A transformation via receipt protocols

To achieve the transformation of Theorem 5.1, we introduce the following notion of a *receipt* protocol, which is essentially a 2-round message authentication protocol in which the party who speaks first chooses the message. Such protocols can be defined via a single function $\mathsf{Receipt}$, as follows.

Definition 5.2. A (k, ℓ, ε)*-receipt protocol (for messages of length d)* is a function $\mathsf{Receipt} : \{0,1\}^d \times \{0,1\}^r \times \{0,1\}^n \to \{0,1\}^\ell$ that satisfies the following: for $Y \equiv U_r$, every $\mu \in \{0,1\}^d$, every X such that $\mathbf{H}_\infty(X|E) \geq k$, and every $\mu' \neq \mu, Y'$ chosen by an adversary given μ, Y, E,

$$\mathbf{H}_\infty(\mathsf{Receipt}(\mu, Y, X) \mid Y, \mathsf{Receipt}(\mu', Y', X)) \geq \log(1/\varepsilon).$$

Given a function $\mathsf{Receipt}$ satisfying this definition, one can construct a protocol as depicted in Protocol 7. The following is immediate.

Theorem 5.3. *Let $\mathsf{Receipt}$ be a function defining a (k, ℓ, ε)-receipt protocol, and assume $\mathbf{H}_\infty(X|E) \geq k$. Then in Protocol 7, Alice accepts with probability $\leq \varepsilon$ if $\mu' \neq \mu$ and accepts with probability 1 if Eve is passive.*

Alice: X	Eve: E	Bob: X
Choose message $\mu \in \{0,1\}^d$ Sample random Y		
	$(\mu, Y) \longrightarrow (\mu', Y')$	
		Compute $T' := \mathsf{Receipt}(\mu', Y', X)$
	$T \longleftarrow T'$	
If $T \neq \mathsf{Receipt}(\mu, Y, X)$ *reject*		

Protocol 7: A receipt protocol

Note that Protocol 7 achieves residual entropy $\geq k - \ell$, because only Bob's message depends on the weak source X.

Besides potentially being of independent interest, receipt protocols are useful because we can give a transformation that increases their residual entropy, which we do not know how to do directly for privacy amplification protocols. Specifically, we prove the following theorem in section 5.2.

Theorem 5.4. *Assume that there exists a polynomial-time (k, ℓ, ε)-receipt protocol for d-bit messages such that Alice communicates $\leq \ell$ bits and $2^{-C\ell} \leq \varepsilon \leq 1/(C\ell)$ for sufficiently large C.*

Then for any $r \leq \log(1/\varepsilon)/100$, there exists a polynomial-time $(k, r, 2^{-\Omega(r)})$-receipt protocol for d-bit messages where Alice communicates $O(\ell)$ bits.

Note that the loss in residual entropy r of the latter receipt protocol is linear in its security parameter, and in particular for some $\varepsilon' = \varepsilon^{\Omega(1)}$ we can obtain a $(k, O(\log(1/\varepsilon')), \varepsilon')$-receipt protocol.

We now show that privacy amplification protocols can be constructed from receipt protocols, and vice versa. In combination with Theorem 5.4, this will prove Theorem 5.1.

One direction, that post-application robust privacy amplification protocols imply receipt protocols, is straightforward. Specifically, Alice and Bob can use the derived key of length m and a MAC with tag length $m/2$ to construct an iMAC protocol (cf. section 2.1) for messages of length d in which Alice speaks first and Bob chooses the message [DW09, Thm. 21]. Using the MAC of Theorem 2.4, this requires only that $d \leq \varepsilon 2^{m/2}$ to bound the attacker's success probability by $O(\varepsilon)$. Then, an iMAC protocol immediately gives a receipt protocol with the same parameters by taking $\mathsf{Receipt}$ to be the function that computes Bob's tag, because the iMAC protocol is secure in particular for a message that Alice chooses and sends to Bob. In summary, we have the following.

Theorem 5.5. *Assume that there exists a polynomial-time 2-round (k, m, ε)-secure privacy amplification protocol with post-application robustness and communication complexity c.*

Then there exists a polynomial-time $(k, c + m/2, O(\varepsilon))$-receipt protocol for messages of length $d \leq \varepsilon 2^{m/2}$ in which Alice communicates $\leq c + d$ bits.

The other direction, that receipt protocols imply (pre- or post-application robust) privacy amplification protocols is slightly more involved, and is depicted in Protocol 8. The idea is that in the receipt protocol, Alice chooses her message S uniformly at random. Bob sends the receipt for S, and uses S to extract a key K with which he authenticates a uniformly random seed W to Alice. Finally, Alice and Bob use W to extract the final key which is the output of the privacy amplification protocol.

Alice: X	Eve: E	Bob: X
Sample random S, Y		
	$(S, Y) \longrightarrow (S', Y')$	
		Compute $T_1' := \mathsf{Receipt}(S', Y', X)$ Compute $K' := \mathsf{Ext}_1(X, S')$ Sample random W' Compute $T_2' := \mathsf{MAC}_{K'}(W')$ Set final $R_B := \mathsf{Ext}_2(X, W')$
	$(T_1, T_2, W) \longleftarrow (T_1', T_2', W')$	
If $T_1 \neq \mathsf{Receipt}(S, Y, X)$ *reject* Compute $K := \mathsf{Ext}_1(X, S)$ **If** $T_2 \neq \mathsf{MAC}_K(W)$ *reject* Set final $R_A := \mathsf{Ext}_2(X, W)$		

Protocol 8: A 2-round privacy amplification protocol using $\mathsf{Receipt}$

Theorem 5.6. *Let* $\mathsf{Receipt}$ *be a (k, ℓ, ε)-receipt for d-bit messages,* $\mathsf{Ext}_1 : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{4\ell}$ *be a $(k, 2^{-4\ell})$-extractor,* $\mathsf{Ext}_2 : \{0,1\}^n \times \{0,1\}^{d'} \to \{0,1\}^{m'}$ *be a $(k - \ell - d, \varepsilon')$-extractor, and* $\mathsf{MAC}_K : \{0,1\}^{d'} \to \{0,1\}^{2\ell}$ *be a MAC with key length $|K| = 4\ell$.*

Then for $k' \geq k + O(\ell)$, Protocol 8 is a (k', m', ε')-secure privacy amplification protocol with residual entropy $k - \ell$, with $\varepsilon' = \varepsilon + O(2^{-\ell})$ and $m' = k' - O(\ell)$ for pre-application robustness or $m' = k' - k - O(\ell)$ for post-application robustness.

Proof. To prove the pre-application robustness of Protocol 8, we consider two cases. If Eve changes S (so $S' \neq S$), then Alice accepts only with probability $\leq \varepsilon$ by the security of the receipt protocol. Note that in attempting to forge $\mathsf{Receipt}(S, Y, X)$ Eve now has an additional 2ℓ bits of information about X given by T_2, but this only decreases the entropy of X from k' to $k' - 2\ell \geq k$.

If instead Eve does not change S, then the security follows from that of MAC given by Theorem 2.4. Specifically, in this case we have $K = K'$, and further $K \approx_{2^{-4\ell}} U_{4\ell}$ even conditioned on S. Note that Eve also has another $|T_1'| = \ell$ bits of information that depend on K, but by Lemma 2.1 and the fact that $K \approx_{2^{-4\ell}} U_{4\ell}$ we have $\mathbf{H}_\infty(K|T_1') \geq \mathbf{H}_\infty(K) - \ell \geq 3\ell - 1$. Thus by Theorem 2.4, Eve can only forge $\mathsf{MAC}_K(W)$ with probability at most

$$\left\lceil \frac{d'}{2\ell} \right\rceil \cdot 2^{2\ell - \mathbf{H}_\infty(K|T_1')} \leq O(2^{-\ell}).$$

Note that here we are assuming $W \neq W'$, as otherwise $R_A = R_B$ and thus Eve would not break robustness.

For post-application robustness the argument is similar. The only difference is that now Eve sees R_B before modifying (T_1, T_2, W), so in order to preserve the security of MAC we require that the length $m' = |R_B|$ of this extra information is $\leq k' - k - O(\ell)$.

To prove the extraction property (assuming by robustness that Eve does not change W from W'), first note that X still has entropy $k - \ell$ even conditioned on (T_1', T_2'), and further $W = W'$ is uniform. However, W and X are now dependent because of T_2'. Following [DW09, Thm. 20] we break this dependence by also giving Eve the key K', which decreases the entropy of X to $k - \ell - d$ but conditioned on which X and W are now independent. Thus by the property of Ext_2, $R_A = R_B$ is ε'-close to uniform conditioned on Eve's view. □

Proof of Theorem 5.1. Let a 2-round (k, m, ε)-secure privacy amplification protocol with post-application robustness be given, and let c denote its communication complexity. By Theorem 5.5, this gives a $(k, c + m/2, O(\varepsilon))$-receipt protocol for messages of length $d \leq \varepsilon 2^{m/2}$ in which Alice communicates $\leq c + d$ bits. Then applying Theorem 5.4, we obtain a (k, ℓ, ε')-receipt protocol for messages of length d where $\varepsilon' = \varepsilon^{\Omega(1)}$ and $\ell = O(\log(1/\varepsilon'))$, provided that $d \leq m/2$. Finally applying Theorem 5.6 gives a privacy amplification protocol with residual entropy $k' - O(\log(1/\varepsilon'))$ and the stated parameters, provided that we can instantiate Ext_1 and Ext_2.

By [GUV09, Thm. 5.12], Ext_1 exists with $d = O(\log n + \ell)$ provided that $4\ell < 0.99k$, which is guaranteed by $k \geq C \log(1/\varepsilon)$ for sufficiently large C. Note that since ℓ and $\log n$ are each $O(\log(1/\varepsilon))$, we can take $d = O(\log(1/\varepsilon)) \leq m/2$. By [GUV09, Thm. 5.14], Ext_2 exists with $d' = \log n + O(\log^2((k - \ell - d)/\varepsilon'))$ provided that $m' \leq k - \ell - d - O(\log(1/\varepsilon'))$, so we can take $m' \geq k' - O(\log(1/\varepsilon'))$ as stated. □

5.2 Increasing residual entropy of receipts

We now prove Theorem 5.4. The transformation that we use employs Raz's strong two-source extractor [Raz05], stated in Theorem 5.7. (Our use of Raz's extractor is inspired by a recent construction of Li [Li12b].)

Theorem 5.7 ([Raz05])**.** *Let S be a (t, e)-source and T an independent (ℓ, e')-source satisfying the following for some $0 < \delta < 1/2$.*

- $t \geq 6 \log t + 2 \log \ell$
- $e \geq (0.5 + \delta)t + 3 \log t + \log \ell$

- $e' \geq 5\log(t-e)$

Then for every $r \leq \delta \cdot \min(t/8,\ e'/40) - 1$*, there exists an explicit efficient function* $\mathsf{Raz} : \{0,1\}^\ell \times \{0,1\}^t \to \{0,1\}^r$ *such that*

$$\Delta(\mathsf{Raz}(T,S),\ U_r \mid S) \leq 2^{-1.5r}.$$

Proof of Theorem 5.4. Let $\mathsf{Receipt}$ be a function defining a (k,ℓ,ε)-receipt protocol. Define $t := 50\ell$, $e' := \log(1/\varepsilon)$, and $e := t - 2(d+|Y|) - r \geq t - 2\ell - r$. Let Raz be the function given by Theorem 5.7 for t, ℓ, e, e', r.

We define a new receipt protocol in Protocol 9.

Alice: X	Eve: E	Bob: X
Choose message $\mu \in \{0,1\}^d$ Sample random Y, S		
	$(\mu, Y, S) \longrightarrow (\mu', Y', S')$	
		Compute $T' := \mathsf{Receipt}(\mu', Y', X)$ Compute $R' := \mathsf{Raz}(T', S')$
	$R \longleftarrow R'$	
Compute $T := \mathsf{Receipt}(\mu, Y, X)$ **If** $R \neq \mathsf{Raz}(T,S)$ *reject*		

Protocol 9: The transformed receipt protocol

To show that this is a $(k, r, 2^{-\Omega(r)})$-receipt protocol, fix any deterministic adversary Eve and consider the following alternate sampling of the random variables appearing in the protocol.

First, sample together μ, μ', Y, and Y' according to their marginal distribution (we assume $\mu \neq \mu'$). After this sampling, we have the following.

- By Lemma 2.1, with probability $\geq 1 - 2^{-r}$ over the choice of (μ, μ', Y, Y') we have

$$\mathbf{H}_\infty(S \mid \mu, \mu', Y, Y') \geq t - 2(d + |Y|) - r = e.$$

- T and T' are now deterministic functions of X.
- X and S are (still) independent, conditioned on the sampling of (μ, μ', Y, Y').

Next, sample T' according to its marginal distribution induced by the sampling of (μ, μ', Y, Y') above. We have now sampled all variables from the original protocol except $T = \mathsf{Receipt}(\mu, Y, X)$. Note that by the security of the original protocol, we have $\mathbf{H}_\infty(T \mid T', \mu, \mu', Y, Y') \geq \log(1/\varepsilon) = e'$. Further, since before we sampled T' it was the case that T and T' were deterministic functions of X and that X and S were independent, fixing T' does not affect the independence of T and S, i.e. T and S remain independent even conditioned on the choice of T'. Lastly note that we still have that $\mathbf{H}_\infty(S \mid T', \mu, \mu', Y, Y') \geq e$.

Thus by Theorem 5.7 and the above analysis, we have

$$\Delta((\mathsf{Raz}(T,S), S),\ (U_r, S) \mid T', \mu, \mu', Y, Y') \leq 2^{-1.5r}.$$

Finally, because S' and R' are deterministic functions of S conditioned on the variables sampled so far, we have

$$\Delta((\mathsf{Raz}(T,S), S, S', R'),\ (U_r, S, S', R') \mid T', \mu, \mu', Y, Y') \leq 2^{-1.5r}.$$

Therefore, Eve can only guess the correct value of $\mathsf{Raz}(T,S)$ with probability $\leq 2^{-r} + (2^{-r} + 2^{-1.5r}) = 2^{-\Omega(r)}$, which completes the proof. $\square$

Finally, we obtain the following corollary by instantiating Theorem 5.1 using the 2-round privacy amplification protocol with post-application robustness due to Dodis and Wichs [DW09, Cor. 4], which requires $k = \Omega(\log^2(1/\varepsilon))$.

Corollary 5.8. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time 2-round (k, m, ε)-secure privacy amplification protocol with post-application robustness that achieves $m = \Omega(\log(1/\varepsilon))$ and residual entropy $k - O(\log(1/\varepsilon))$.*

6 Applications to the Bounded Retrieval Model

In the Bounded Retrieval Model (BRM) [CLW06, Dzi06], Alice and Bob share an (intentionally) very large secret key X. The idea is that the size of X makes it infeasible for an attacker Eve to learn the entire string, even if she has infiltrated either Alice or Bob's storage device, because of limits on the amount of data that can be transmitted out of the device. Thus as in previous sections we assume that Eve has some adversarially chosen side information E about X, but that $k := \mathbf{H}_\infty(X|E)$ is not too small. Specifically here we think of $k = \alpha n$ for some constant $0 < \alpha < 1$.

Since reading the entire string X would be prohibitively inefficient, any function used by Alice or Bob that takes X as input must only read a small number of positions, i.e. it must be *locally computable*. Dodis and Wichs observe [DW09, Sec. 5] that their privacy amplification protocol, used in Corollary 5.8 above, has the property that each function taking X as input is a standard extractor (as opposed to non-malleable — recall that the protocol uses *look-ahead extractors*, which are constructed from standard extractors). These can be replaced with the constructions of locally computable extractors due to Vadhan [Vad04], and thus the protocol works even in the BRM.

One downside of the [DW09] protocol is that the second message (which depends on X) has length $\Omega(\log^2(1/\varepsilon))$, and thus the loss in residual entropy is $\Omega(\log^2(1/\varepsilon)) = \Omega(m^2)$. It would be more desirable to have loss in residual entropy $O(m)$, as then Alice and Bob could derive a total of $\Omega(k/m)$ secret keys from the weak source X, as opposed to only $O(k/m^2)$ keys. (Deriving many short keys from a single long-term secret is a prominent use case for the BRM; note that post-application robustness is therefore the right notion of security in this setting.)

Corollary 5.8 shows that the loss in residual entropy can be reduced to $O(m)$, allowing Alice and Bob to derive $\Omega(k/m)$ keys which is optimal up to constant factors. This protocol remains locally computable and thus applicable to the BRM, because still every function that takes X as input is a standard extractor and can be replaced by a locally computable extractor. Specifically, each such function is either an extractor from the [DW09] protocol, or one arising from the transformation of Theorem 5.6. In summary, we have the following.

Theorem 6.1. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time 2-round $(k, m = \Omega(\log(1/\varepsilon)), \varepsilon)$-secure privacy amplification protocol in the BRM with post-application robustness and residual entropy $k - O(\log(1/\varepsilon))$, thus allowing a total of $\Omega(k/m)$ keys to be derived.*

At the expense of moving from two to four rounds, we can obtain a BRM protocol that additionally has *source privacy* by instead plugging the [DW09, Cor. 4] protocol into the transformation of Theorem 3.12. The resulting protocol also has optimal residual entropy because the final message of length $O(\log(1/\varepsilon))$ is the only one depending on X, and is "BRM friendly" because again the only functions that touch X are standard extractors. (Note that Bob's computation in the [DW09, Cor. 4] protocol has the two-stage form depicted in Protocol 4, and thus Theorem 3.12 is applicable.)

Theorem 6.2. *For $k = \Omega(\log^2(1/\varepsilon))$, there exists an explicit polynomial-time 4-round $(k, m = \Omega(\log(1/\varepsilon)), \varepsilon)$-secure (k, ε)-private privacy amplification protocol in the BRM with post-application robustness and residual entropy $k - O(\log(1/\varepsilon))$, thus allowing a total of $\Omega(k/m)$ keys to be derived.*

References

[BBCM95] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, 1995.

[BBR88] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, April 1988.

[BF11] Niek J. Bouman and Serge Fehr. Secure authentication from a weak key, without leaking information. In *EUROCRYPT*, pages 246–265, 2011.

[CKOR10] Nishanth Chandran, Bhavana Kanukurthi, Rafail Ostrovsky, and Leonid Reyzin. Privacy amplification with asymptotically optimal entropy loss. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, 2010.

[CLW06] Giovanni Di Crescenzo, Richard J. Lipton, and Shabsi Walfish. Perfectly secure password protocols in the bounded retrieval model. In *TCC*, pages 225–244, 2006.

[CRS12] Gil Cohen, Ran Raz, and Gil Segev. Non-malleable extractors with short seeds and applications to privacy amplification. In *IEEE Conference on Computational Complexity*, pages 298–308, 2012.

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

[DKK$^+$12] Yevgeniy Dodis, Bhavana Kanukurthi, Jonathan Katz, Leonid Reyzin, and Adam Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. *IEEE Transactions on Information Theory*, 58(9):6207–6222, 2012.

[DKSS09] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to kakeya sets and mergers. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009.

[DLWZ11] Yevgeniy Dodis, Xin Li, Trevor D. Wooley, and David Zuckerman. Privacy amplification and non-malleable extractors via character sums. In *FOCS*, pages 668–677, 2011.

[DORS08] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38:97–139, 2008.

[DS02] Yevgeniy Dodis and Joel Spencer. On the (non)universality of the one-time pad. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, page 376. IEEE Computer Society, 2002.

[DS05] Yevgeniy Dodis and Adam Smith. Entropic security and the encryption of high entropy messages. In *TCC*, pages 556 – 577, 2005.

[DW09] Y. Dodis and D. Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, 2009.

[DY13] Yevgeniy Dodis and Yu Yu. Overcoming weak expectations. In *TCC*, pages 1–22, 2013.

[Dzi06] Stefan Dziembowski. Intrusion-resilience via the bounded-storage model. In *TCC*, pages 207–224, 2006.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *Journal of the ACM*, 56(4), 2009.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[KR09] B. Kanukurthi and L. Reyzin. Key agreement from close secrets over unsecured channels. In *EUROCRYPT*, pages 206–223, 2009.

[Li12a] Xin Li. Design extractors, non-malleable condensers and privacy amplification. In *STOC*, pages 837–854, 2012.

[Li12b] Xin Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. *CoRR*, abs/1211.0651, 2012.

[Li12c] Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. In *FOCS*, pages 688–697, 2012.

[Lu04] Chi-Jen Lu. Encryption against storage-bounded adversaries from on-line strong extractors. *J. Cryptology*, 17(1):27–42, 2004.

[Mau92] Ueli M. Maurer. Protocols for secret key agreement by public discussion based on common information. In *CRYPTO*, pages 461–470, 1992.

[MW97] Ueli M. Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *CRYPTO '97*, 1997.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996.

[Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.*, 13(1):2–24, 2000.

[RW03] R. Renner and S. Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In *CRYPTO*, pages 78–95, 2003.

[Sti94] Douglas Stinson. Universal hashing and authentication codes. *Designs, Codes and Cryptography*, 4(3):369–380, 1994.

[Vad04] Salil P. Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *J. Cryptology*, 17(1):43–77, 2004.