Optimal Randomizer Efficiency in the Bounded-Storage Model*

Stefan Dziembowski**
Institute of Informatics, University of Warsaw
Banacha 2, PL-02-097 Warsaw, Poland
std@mimuw.edu.pl

Ueli Maurer***
Department of Computer Science, ETH Zurich
CH-8092 Zurich, Switzerland
maurer@inf.ethz.ch

Abstract. In the bounded-storage model for information-theoretically secure encryption and key-agreement one can prove the security of a cipher based on the sole assumption that the adversary's storage capacity is bounded, say by s bits, even if her computational power is unlimited. Assume that a random t-bit string R is either publicly available (e.g. the signal of a deep space radio source) or broadcast by one of the legitimate parties. If $s < t$, the adversary can store only partial information about R. The legitimate sender Alice and receiver Bob, sharing a short secret key K initially, can therefore potentially generate a very long n-bit one-time pad X with $n \gg |K|$ about which the adversary has essentially no information.
All previous results in the bounded-storage model were partial or far from optimal, for one of the following reasons: either the secret key K had to be longer than the derived one-time pad ($n < |K|$), or t had to be extremely large ($t > ns$), or the adversary was assumed to be able to store only s actual bits of R rather than arbitrary s bits of information about R, or the adversary received a non-negligible amount of information about X.
In this paper we prove the first non-restricted security result in the bounded-storage model: K is short, X is very long, and t needs to be only moderately larger than $s + n$. In fact, s/t can be arbitrarily close to 1 and hence the storage bound is essentially optimal. The security can be proved also if R is not uniformly random, provided that the min-entropy of R is sufficiently greater than s.

Keywords. Bounded-storage model, unconditional security, one-time pad, information theory, min-entropy.

1 Introduction

In view of the growing dependence of the information society on cryptography, security proofs for cryptographic schemes are of great importance. Some of the major achievements of the past decade or two in research in cryptography are precise security definitions for many types of cryptographic schemes, as well as security proofs for a number of proposed schemes, relative to these definitions and various assumptions, including typically the assumption that a particular computational problem (e.g. factoring integers) is intractable.

The security of every cryptographic system depends on certain assumptions. A main goal of research in cryptography is to reduce the assumptions underlying a security proof. Usually, not all assumptions are stated explicitly. For instance, two obvious such assumptions are (1)

* The conference version of this paper appeared as [13].
** Part of this work was done while the first author was a Ph.D. student at BRICS, Denmark and a Post-Doc at ETH Zurich, Switzerland. Supported in part by the Polish KBN grant no. 7 T11C 027 20 and by the Foundation for Polish Science (FNP).
*** Supported in part by the Swiss National Science Foundation.

that randomness exists, i.e., that one can generate random keys, and (2) that such keys are independent of an adversary's view. More generally, one assumes that it is impossible to read somebody's mind, at least not in a cryptographically relevant context.

Almost all cryptographic systems in practical use are based on the two further assumptions that (3) the adversary's computational resources are bounded, and (4) that a certain computational problem is hard, i.e., requires an infeasible amount of time to solve, given the assumed upper bound on the adversary's resources. Assumption (4) could potentially be dropped if a complexity-theoretic lower bound could be proved for the problem at hand, but such lower bound proofs appear to be far beyond the reach of known techniques in complexity theory, even for the classical Turing machine model, let alone for more general models consistent with the laws of physics, like a quantum computer.

In contrast, information-theoretically secure systems rely on neither of the assumptions (3) or (4), i.e., the adversary is assumed to have unbounded computing power. However, the security of such a system may rely on an assumption about the probabilistic behavior of Nature, for instance of a noisy channel [19] or a quantum measurement [4].

This paper is concerned with information-theoretically secure encryption and key-agreement. More precisely, we consider the secure expansion of a short shared secret key into a very long shared secret key. Using the one-time pad encryption method, a key-agreement scheme can directly be converted into an encryption scheme. If the one-time pad (i.e., the key) is essentially uniformly distributed, then the one-time pad is essentially perfect. We can therefore concentrate on key expansion, i.e., on generating the one-time pad.

Such key expansion at first glance apparently contradicts the known bounds on the key size of a perfect or close to perfect cipher. Shannon's classical result [23] states that if the adversary Eve has full access to the (one-way) communication channel from Alice to Bob and is missing only the secret key, then the entropy of the secret key is bounded from below by the entropy of the message to be securely transmitted from Alice to Bob. It was proved in [19] that this bound holds also in the more natural and practically relevant scenario where Alice and Bob can communicate arbitrarily over an insecure (but even authenticated) channel accessible to Eve.[1] The motivation of the bounded-storage model is to overcome this impossibility result by a twist in the model, namely by introducing an additional random string too large to be stored completely by the adversary.

2 The Bounded-Storage Model and Related Work

2.1 Motivation

The two main parameters specifying the adversary Eve's resources are her computing power (e.g. specified in MIPS[2]) and her storage capacity (e.g. specified in terabytes). Complexity-theoretic cryptography is based on an assumed upper bound on Eve's computing power (and possibly storage capacity). The natural idea of the bounded-storage model, proposed in [18], is that one makes a sole (conservative) assumption about Eve's storage capacity (e.g. that it is at most one petabit, i.e. 10^{15} bits), but no assumption whatsoever about her computing power. Let s be the assumed bound on Eve's storage capacity (in bits).

Ciphers in the bounded-storage model make use of a very large amount of auxiliary information, denoted R and called public randomness or simply the *randomizer*. The randomizer

[1] This implies that there is no information-theoretically secure version of two-way public-key cryptography.

[2] Note that MIPS is not a precisely-defined unit of computation. For a concrete security proof the computing power would have to be specified precisely.

R could for instance be a random bit sequence broadcast by a satellite or transmitted between the legitimate parties, or the signal of a deep-space radio source. If R is a t-bit random string, then $t > s$ is required to guarantee that Eve cannot store R completely. More generally, it suffices to assume a bound on the min-entropy of R.[3] The restriction $t > s$ immediately shows the inherent (but only) limitation of the bounded-storage model: In order to be realistic, s and hence also the size t of R must be very large. Nevertheless, schemes based on this model for which t is not much larger than s may be on the verge of being practical, even for very powerful adversaries. The main challenge, which we solve in this paper, is to devise a provably secure scheme with s close to t.

Let us comment briefly on the practicality of the bounded-storage model, but in this paper we do not give a detailed feasibility analysis for current technology. A recent article in the New York Times (Feb. 20, 2001) and other media reports suggested that such schemes may be used in practice. However, because of the inherent condition $s < t$, the feasibility depends on possible advances in storage and in communication technology (see [12] or [11], Section 1.3.1 for an analysis of the current technology).

As a concrete example of what is proved in this paper (Example 10), assume that the adversary's storage capacity is at most one petabit, (i.e. $s = 10^{15}$), that Alice and Bob share a 6000-bit secret key and that they (and Eve) have access to a random source emitting 100 gigabits (10^{11} bits) per second, which they access for about one day and a half, i.e., $t = 1.25 \cdot 10^{16}$. Then they can derive an expanded key of length 10 gigabits (i.e. $n = 10^{10}$) about which the adversary has essentially no information, even if she uses an optimal strategy. Alice and Bob need to read only $1.25 \cdot 10^{12}$ bits from the random source.

2.2 From Book Ciphers to the Bounded-Storage Model

One can view book ciphers, known from spy stories, as a special case of such a randomized cipher. Assume that Alice and Bob agree, not necessarily secretly, on a particular book of which they each have a copy. The book plays the role of the randomizer. To use the book cipher, Alice and Bob agree on a secret key consisting of a page number and a pointer to a letter on that page. The text following that letter is used as a one-time pad (modulo 26) to encrypt a (single) message. It is clear that if Eve also has a copy of the book and knows a sufficiently long ciphertext, she can find the key by an exhaustive key search, provided the plaintext is redundant. In pre-OCR times this was a very cumbersome task, but not infeasible because the size of the key space corresponds only to the length of the text in the book. It is obvious how a binary version involving a long random string R instead of a book would work: plaintext and key are binary sequences and encryption is the bit-wise XOR operation. This is the basic idea underlying Massey and Ingemarsson's so-called Rip van Winkle cipher [17].

In our context, however, because R is not a book but rather an immensely long bit string with $|R| = t$, it is realistic to assume that Eve does not know the entire value of R but has stored s bits of information about R. If, for example, $s = t/2$, it is clear that Eve could for the discussed binary version of the book cipher obtain on average about half of the information about the plaintext, which would be completely insecure. This problem can be solved by encrypting the plaintext with several (say m) independent keys (but the same randomizer), i.e., by defining the one-time pad as the (bit-wise modulo 2) sum of m blocks of R beginning at independent random locations within R, where the key consists of the m starting points. This, in essence, is the scheme proposed in [18], which we also use here and which was also

[3] It was first proposed by Chor and Goldreich [8] to measure the non-uniformity of a random variable by its min-entropy.

used in [2], but for a size of the one-time pad of only $n = 1$ bits. The main difference between the schemes is that in [18] each of the m blocks is taken from a separate non-overlapping part of R, with a cyclic continuation if the block reaches the end of the part, whereas in this paper no cyclic extension is used and in [2] the non-blocked scheme sketched above is used. In the subsequent papers of Lu [16] and Vadhan [25], more complicated constructions with a reduced key size are proposed.

2.3 Definition of the Bounded-Storage Model

We now define the bounded-storage model for key-expansion (and encryption) more formally. Alice and Bob share a short secret *initial key* K, selected uniformly at random from a key space $\mathcal{K}$, and they wish to generate a much longer n-bit *expanded key* $X = (X_1, \ldots, X_n)$ (i.e. $n \gg \log_2 |\mathcal{K}|$).

In a first phase, a t-bit string R is available to all parties, i.e., the randomizer space is $\mathcal{R} = \{0,1\}^t$. For instance, R is sent from Alice to Bob or broadcast by a satellite. Rather than assuming that R is uniformly random, which may be unrealistic, one can also simply assume a lower bound on the min-entropy $H_\infty(R)$ of R. Alice and Bob apply a known *key-expansion function* $f : \mathcal{R} \times \mathcal{K} \to \{0,1\}^n$ to compute the expanded key as $X = f(R, K)$. Of course, the function f must be efficiently computable and based on only a very small portion of the bits of R such that Alice and Bob need not read the entire string R.

Eve can store arbitrary s bits of information about R, i.e., she can apply an arbitrary storage function $h : \mathcal{R} \to \mathcal{U}$ for some $\mathcal{U}$ with the only restriction that $|\mathcal{U}| \leq 2^s$.[4] The memory size during the evaluation of h need not be bounded. The value stored by Eve is $U = h(R)$. After storing U, Eve loses the ability to access R. All she knows about R is U. In order to prove as strong a result as possible, we assume that Eve can now even learn K, although in a practical system one would of course keep K secret.

A key-expansion function (or cipher) f is secure in the bounded-storage model if, with overwhelming probability, Eve, knowing U and K, has essentially no information about X. More precisely, the conditional probability distribution $P_{X|U=u,K=k}$ is very close to the uniform distribution over the n-bit strings, with overwhelming probability over values u and k. Hence X can be used as a secure one-time pad (cf. Section 3.3). Obviously, this is not possible for $s \geq t$ nor for $s \geq H_\infty(R)$ (without making a further assumption about R), but it should hold for as large a storage bound s as possible, ideally for $s = \nu H_\infty(R)$ for ν close to 1. The ratio

$$\nu := s/H_\infty(R).$$

will be called the *randomness efficiency* of a scheme. Actually, ν may depend on the ratio $H_\infty(R)/t$, but for uniform R we have $H_\infty(R) = t$ and thus $\nu = s/t$.

2.4 Previous Results for the Bounded-Storage Model

The bounded-storage model was introduced by Maurer in 1990 [18]. However, the proposed cipher was proved secure in [18] only under the assumption that Eve stores s *actual* bits of R rather than the result of an arbitrary function applied to R.[5] The s bits of R can be

[4] Since for every probabilistic strategy there is a best choice of the randomness, we can without loss of generality consider only deterministic adversary strategies.

[5] This was justified by considering the following scenario. The t-bit randomizer R is assumed to be permanently accessible to all parties, but it is too long to be read entirely. The basic operation is that of reading a bit of R. As a somewhat unrealistic but illustrative example it was proposed to use the surface structure of

accessed using an arbitrary adaptive strategy, where the position of each new bit depends on the previously seen bits. The scheme is secure for, say, $s \leq t/2$.[6]

As discussed above, in the scheme of [18], R is divided into m parts of l bits, i.e. $t = lm$, and each bit of X is the XOR of m bits, one from each part. The key K determines which bits are XOR-ed. We refer to Section 4.1 for a precise description of the scheme analyzed in this paper, which is essentially the same as that of [18]. The main problem left open in [18] and solved here is to prove the security of this scheme in the model where the adversary is allowed to compute an arbitrary function of the random string (as described in Section 2.3).

Cachin and Maurer [7] proposed a scheme in which Eve is allowed to access arbitrary s bits of information about R, but the probability that Eve can obtain a non-negligible amount of information about X is non-negligible (e.g. 0.0001). Another scheme proposed there requires no secret key K but is impractical.

A major step towards solving the open problem of [18] was achieved by Aumann, Ding, and Rabin [2, 1, 11, 12], using a scheme very similar to that of [18]. The core technical argument is a security proof for a scheme for generating a single key bit (i.e. $n = 1$) and for randomness efficiency $\nu_{\mathrm{ADR}} \approx 0.2$. Of course, in practice one is interested in deriving a key of length $n > 1$. In order to achieve this goal, one can use the single-bit scheme as a building block. This can be done in two different, but in a sense dual ways.

- Execute the single-bit scheme n times [2, 1, 11] with the same key, but with independently chosen randomizers. The drawback of this approach is that the security can be proven only if $s \leq \nu_{\mathrm{ADR}} \cdot t/n$, i.e., the randomness efficiency $\nu = \nu_{\mathrm{ADR}}/n$ decreases inversely proportional to n.
- Execute the single-bit scheme n times [1, 11, 12] with the same randomizer, but independently chosen initial keys. Here the security can be proved assuming that $s \leq \nu_{\mathrm{ADR}} \cdot t$, but the drawback is that in order to derive an n-bit key one needs an initial key much longer than n. The reason why this result is nevertheless of interest is that the same key can be used many times with new randomizers.

The bounded-storage model was also studied in the context of secure two-party computations [11, 10, 6].

2.5 Contributions of this Paper

The main open question of [18, 2, 1] is whether significant key-expansion (i.e., $n \gg \log_2 |\mathcal{K}|$) with constant randomness efficiency ν is possible. We solve this problem which has both theoretical and possibly practical implications. The technical contributions of the paper are divided into two parts. The first part (Sections 4 and 5) addresses parameter sizes that are closest to being of practical interest. Theorem 8 and Corollary 9 state that for reasonable parameter sizes, secure key-expansion is possible for ν in the 10% range. In the second part (Section 6) we prove, as a theoretical result, that ν can be arbitrarily close to 1.

Let $w := \log_2 |\mathcal{K}|$ be the length of the initial key. Our results imply, for example, that for every $\nu < 1$ and $z > 3$, and any key expansion ratio $c(m)$ with $1 < c(m) = 2^{o(m)}$, there exists a family of schemes (with parameter m) such that $t = O(m^z)$, $w = O(m \log m)$, and

the moon as a huge array of random bits. The scheme is secure even if the adversary can adaptively choose the bit positions to access and even when she has arbitrary prior knowledge about the plaintext.

[6] In fact, the cipher is perfect with overwhelming probability, i.e., with overwhelming probability Eve gets no information whatsoever about X (while with negligible probability she may learn something about X). This statement is slightly stronger than the statement that Eve gets only a negligible amount of information.

$n = c(m)w + o(w)$, and such that the deviation from uniform of the expanded key, from the adversary's viewpoint, decreases exponentially in m, provided that $s \leq \nu t$.

The scheme we propose is essentially the same as that used in previous works on the bounded-storage model. The main contribution of this paper is therefore to provide a new, stronger security proof. Some main steps of the proof are as follows. First, we observe that it suffices to prove that the adversary cannot predict well the last bit of the expanded key when given all other bits of the expanded key. Second, we prove that for any strategy of the adversary, the fraction of randomizers for which his strategy gives him a non-negligible advantage in the prediction problem, is negligible. In this step, we need to make use of Azuma's inequality to cope with statistical dependencies that prevent us from using the Chernoff bound.

2.6 Subsequent Results

After the presentation of these results in [13], but independently, Lu [16] pointed out that the construction of a secure cryptosystem in the bounded-storage model corresponds to the design of a so-called *randomness extractor* (first defined in [21], see also [22, 24]) with a special property, which he called an *on-line* extractor. An extractor E takes two inputs, say W and Z, and has the property that $E(W, Z)$ is very close to uniform if W has sufficient min-entropy and Z is uniform and independent of W. Actually, in our context one needs a so-called *strong* extractor whose output includes the second input Z.

The extractor argument also shows that the distribution of the randomizer R need not be uniform but must only have sufficient min-entropy. The same is true for our scheme, essentially without modifying our original proof (see [13]) which was stated for the case of uniform R. Actually, our scheme is a (very simple) construction of a strong on-line extractor. Subsequently, Vadhan [25] improved further on the construction of [16] by considering the general problem of designing locally computable extractors. Both these schemes also have optimal randomizer efficiency.

While these subsequent papers have provided insight and improved on the size of the initial key, their schemes appear to be more complicated than ours which can be efficiently implemented with a very simple process for selecting bits from the randomizer and for XOR-accumulating them to obtain the expanded key. The accumulator for the expanded key is the only memory that is required. Simplicity may be significant since, if the bounded-storage model should ever be practical, then the data rates will have to be at the very limits of the available technology, possibly making irregular bit selection or significant processing of information infeasible. But only future technology will show which type of scheme is most advantageous, and it is an open research problem to define simplicity meaningfully and to find even better schemes.

3 Preliminaries

3.1 Probability-Theoretic Preliminaries

Throughout the paper we will use capital letters to denote random variables and small letters to denote values they can take on.

The min-entropy of a random variable X with probability distribution P_X over $\mathcal{X}$ is defined as

$$H_\infty(X) := -\log_2 \left(\max_{x \in \mathcal{X}} P_X(x) \right)$$

$$= \min_{x \in \mathcal{X}} \left(-\log_2 P_X(x)\right).$$

The deviation of a probability distribution p over an alphabet $\mathcal{U}$ from the uniform distribution can be measured by its statistical distance $d(p)$ from the uniform distribution:

$$d(p) := \tfrac{1}{2} \sum_{u \in \mathcal{U}} \left| p(u) - \frac{1}{|\mathcal{U}|} \right|.$$

For a random variable X we will also write simply $d(X)$ instead of $d(P_X)$. Similarly, for an event A, we define

$$d(X|A) := d(P_{X|A}).$$

The following notation and definitions are motivated by information theory. The distance of X from uniform, given random variable Y, is defined as the expected distance of $P_{X|Y=y}$ from uniform when averaged over values y of Y:

$$\begin{aligned} d(X|Y) &:= \sum_{y \in \mathcal{Y}} P_Y(y)\, d(X|Y=y) \\ &= \sum_{y \in \mathcal{Y}} P_Y(y) \tfrac{1}{2} \sum_{x \in \mathcal{X}} \left| P_{X|Y}(x,y) - \frac{1}{|\mathcal{X}|} \right| \\ &= \tfrac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left| P_{XY}(x,y) - \frac{P_Y(y)}{|\mathcal{X}|} \right|. \end{aligned}$$

It is well-known (and easy to prove) that if X is a binary random variable, then $d(X|Y)$ measures the optimal advantage (above $\frac{1}{2}$) of a predictor g guessing X when given Y:

Lemma 1 *If X is binary, then* $\max_{g\colon \mathcal{Y} \to \mathcal{X}} P\left(g(Y) = X\right) = \frac{1}{2} + d(X|Y)$.

Lists of random variables are abbreviated by simply concatenating their names; for example $d(XY)$ is used instead of $d((X,Y))$.

Lemma 2 $d(XY) \le d(X|Y) + d(Y)$.

Proof.

$$\begin{aligned} d(X|Y) + d(Y) &= \tfrac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left| P_{XY}(x,y) - \frac{P_Y(y)}{|\mathcal{X}|} \right| + \tfrac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left| \frac{P_Y(y)}{|\mathcal{X}|} - \frac{1}{|\mathcal{X}||\mathcal{Y}|} \right| \\ &= \tfrac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left(\left| P_{XY}(x,y) - \frac{P_Y(y)}{|\mathcal{X}|} \right| + \left| \frac{P_Y(y)}{|\mathcal{X}|} - \frac{1}{|\mathcal{X}||\mathcal{Y}|} \right| \right) \\ &\ge \tfrac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left| P_{XY}(x,y) - \frac{1}{|\mathcal{X}||\mathcal{Y}|} \right| = d(XY), \end{aligned}$$

where the inequality follows from the triangle inequality for interval lengths. □

It follows immediately from Lemma 2 that

$$d(X_1 \cdots X_n) \le \sum_{i=1}^{n} d(X_i | X_1 \cdots X_{i-1}) \tag{1}$$

and, more generally, we have:

Lemma 3 $d(X_1 \cdots X_n|Y) \leq \sum_{i=1}^{n} d(X_i|X_1 \cdots X_{i-1}Y)$.

Lemma 4 $d(X|Y) \geq d(X)$.

Proof.

$$\begin{aligned} d(X|Y) &= \tfrac{1}{2} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \left| P_{XY}(x,y) - \frac{P_Y(y)}{|\mathcal{X}|} \right| \\ &\geq \tfrac{1}{2} \sum_{x\in\mathcal{X}} \left| \sum_{y\in\mathcal{Y}} \left(P_{XY}(x,y) - \frac{P_Y(y)}{|\mathcal{X}|} \right) \right| \\ &= \tfrac{1}{2} \sum_{x\in\mathcal{X}} \left| P_X(x) - \frac{1}{|\mathcal{X}|} \right| = d(X). \end{aligned}$$ □

When X is a binary random variable, then Lemma 4 (in view of Lemma 1) states the intuitive and well-known fact that the advantage in (optimally) guessing X cannot decrease when learning another random variable Y. An intuitive proof would be that a guesser that ignores Y is admissible.

The above notation and results carry over naturally to a setting where everything is conditioned on the event $Z = z$ that some random variable Z takes on the value z. For instance,

$$d(X|Y, Z=z) := \sum_{y\in\mathcal{Y}} P_{Y|Z}(y,z)\, d(X|Y=y, Z=z),$$

and Lemma 1 becomes

$$\max_{g:\,\mathcal{Y}\to\mathcal{X}} P\left(g(Y) = X \mid Z=z\right) = \tfrac{1}{2} + d(X|Y, Z=z).$$

Similarly, Lemma 3 becomes

$$d(X_1 \cdots X_n|Y, Z=z) \leq \sum_{i=1}^{n} d(X_i|X_1 \cdots X_{i-1}Y, Z=z).$$

3.2 Martingales

We will need the following result from the theory of martingales and refer to [20] for more details on the subject.

Definition 5 A sequence of real-valued random variables $S_1, \ldots, S_l$ is called a *martingale difference sequence* if, for every j and every $s_1, \ldots, s_{j-1}$,

$$E\left[S_j \mid S_1 = s_1, \ldots, S_{j-1} = s_{j-1}\right] = 0.$$

The following lemma follows directly from Azuma's inequality (see for example Theorem 4.16 in [20], page 92, or [3] for the original result). To derive Lemma 6 from Theorem 4.16 in [20], set $\lambda := \tau l a$ and $X_0 := 0$, and for every $i = 1, \ldots, l$, let $X_j := \sum_{j=1}^{i} S_j$ and $c_i := a$.

Lemma 6 *Let $S_1, \ldots, S_l$ be a martingale difference sequence such that $|S_i| \leq a$ for $1 \leq i \leq l$. Then, for every $\tau > 0$,*

$$P\left(\left| \sum_{j=1}^{l} S_j \right| \geq \tau l a \right) \leq 2e^{-l\tau^2/2}.$$

If the variables $S_1, \ldots, S_l$ are independent, then Lemma 6 follows from the Chernoff bound.

3.3 One-time Pad and Statistical Indistinguishability

The one-time pad encrypts a message M from message space $\mathcal{M} = \{0,1\}^n$ by adding (bit-wise modulo 2) a random key $X \in \mathcal{M}$, resulting in ciphertext $C = M \oplus X$. More generally, one could apply any group operation $\star$ on $\mathcal{M}$, i.e., $C = M \star X$. It is well-known that this generalized one-time pad is perfect (i.e., C is independent of M) if X is uniformly distributed over $\mathcal{M}$.

Since we propose to use the expanded key as the key in a one-time pad encryption, we need to investigate the potential security degradation due to the slight deviation of the key X from the uniform distribution (from the adversary's viewpoint). If X is not uniformly distributed, then the one-time pad is not perfect. The natural relaxation of perfect secrecy is to consider the statistical indistinguishability[7] of the cipher, i.e., the optimal advantage, for any unbounded distinguisher, of distinguishing the ciphertexts of two different messages m and m', maximized over the choice of the pair (m, m').

Lemma 7 *For the one-time pad with key X, the advantage in the statistical indistinguishability definition is upper bounded by $2d(X)$.*

Proof. Let $\delta(P_V, P_W)$ be the distance between the distributions P_V and P_W (over the same alphabet), i.e., $\delta(P_V, P_W) = \frac{1}{2}\sum_v |P_V(v) - P_W(v)|$. Let P_U denote the uniform distribution. Then $d(P_V) = \delta(P_V, P_U)$. Let $C = m \star X$ and $C' = m' \star X$ be the two ciphertexts for messages m and m', respectively. The advantage in the statistical indistinguishability game is

$$\delta(P_C, P_{C'}) \le \delta(P_C, P_U) + \delta(P_{C'}, P_U) = d(C) + d(C') = 2d(X),$$

where we have made use of the triangle inequality for δ. □

This lemma justifies that we restrict our attention to proving that the statistical distance of the expanded key X from uniform is negligible in the adversary's view.

4 The Main Theorem

4.1 Description of the Cipher

The randomizer $R \in \mathcal{R} = \{0,1\}^t$ is interpreted as being arranged in a matrix with m rows, denoted $R(1), \ldots, R(m)$, for some $m \ge 1$ called the *height* of the randomizer. Each row consists of $l + n - 1$ bits, for some $l \ge 1$ called the *width* of the randomizer. Hence $t = m(l + n - 1)$ and R can be viewed as an $m \times (l + n - 1)$ matrix (see Figure 1). The initial key $K = (K_1, \ldots, K_m) \in \mathcal{K} = \{1, \ldots, l\}^m$ selects one starting point within each row, and the expanded key $X = (X_1, \ldots, X_n)$ is the component-wise XOR of the m blocks of length n beginning at these starting points K_i, i.e.,

$$X = f(R, K),$$

where $f : \mathcal{R} \times \mathcal{K} \to \{0,1\}^n$ is defined as follows. For $r \in \mathcal{R}$ and $k = (k_1, \ldots, k_m) \in \mathcal{K}$,

$$f(r,k) := \left(\bigoplus_{i=1}^{m} r(i, k_i), \ldots, \bigoplus_{i=1}^{m} r(i, k_i + n - 1) \right), \tag{2}$$

where $r(i,j)$ denotes the jth bit in the ith row of r. This is illustrated in Figure 1.

[7] This notion was also called *semantic security* of the cipher in [1, 11].

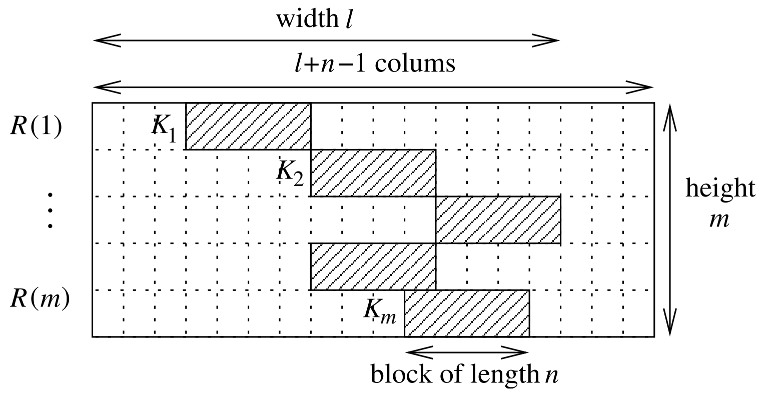

Fig. 1. Illustration of the scheme for deriving an expanded n-bit key $X = (X_1, \ldots, X_n)$, to be used as a one-time pad, from a short secret initial key $K = (K_1, \ldots, K_m)$. The randomizer R is interpreted as a $m \times (l+n-1)$ matrix with rows $R(1), \ldots, R(m)$ of length $l + n - 1$. The expanded key X is the component-wise XOR of m blocks of length n, one selected from each row, where K_i is the starting point of the ith block within the ith row $R(i)$.

The random experiment consists of selecting $K \in \mathcal{K}$ uniformly at random and, independently $R \in \mathcal{R}$ with some (not necessarily uniform) distribution P_R (known to the adversary). All probabilities considered in this paper are for events in this random experiment, unless stated otherwise.

4.2 Statement and Explanation of the Main Theorem

Consider fixed parameters l, m, n and s and remember that $t = m(l+n-1)$. In the following we consider an arbitrary but fixed storage function $h : \mathcal{R} \to \mathcal{U}$ with $|\mathcal{U}| \leq 2^s$. Recall that

$$U = h(R)$$

is the random variable corresponding to the value stored by Eve.

We are interested in proving that from Eve's point of view, when given particular values u for U and k for K, the expanded key X is very close to uniformly distributed, i.e., $d(X|U = u, K = k)$ is very small, with very high probability (over values for U and K). (Remember that we assume that Eve can learn K after she loses the ability to access R.) It suffices to prove that the expected value $d(X|UK)$ of $d(X|U = u, K = k)$ is very small. Markov's inequality (see e.g. [9], page 57) states that for every positive-valued random variable Z and any $\alpha > 0$, $P(Z \geq \alpha) \leq E[Z]/\alpha$. This allows to convert the statement of Theorem 8, namely that $d(X|UK)$ is negligible, into the statement that $d(X|U = u, K = k)$ can exceed a certain (very small) bound only with negligible probability.

The following theorem, the main result of the paper, is proved in Section 4.3. The theorem implies that the key expansion function is actually an extractor.

Theorem 8 *For l, m, n, s, t, R, K and U as defined above and for any $\tau \in [0,1]$ and $\xi \in [0,1]$,*

$$d(X|UK) \leq n \cdot \left(2^{s+t-H_\infty(R)-\Delta} + \epsilon\right), \tag{3}$$

where

$$\Delta := lm(1-\xi)\tau^2 \underbrace{(\log_2 e)/2}_{\geq 0.721} - m(1 + (1-\xi)\,(m \log_2 l + n + 1)) \tag{4}$$

and

$$\epsilon := \tau^{\xi m}/2. \tag{5}$$

The first term in (4) is $O(lm)$ and dominates the second term for $m, n \ll l$. The implication of Theorem 8 is stated below, as an example, for concrete parameters with randomness efficiency $\nu = 0.08$.

Corollary 9 *If R is uniformly distributed and $l, m,$ and n satisfy $m \log_2 l \leq n$ and $l > 100$, then, for $s := 0.08t - 1.5m(n+1)$,*

$$d(X|UK) \leq n2^{-m/2}. \tag{6}$$

The assumption $m \log_2 l \leq n$ reflects the fact that we are interested in key expansion (the length in bits of the initial key is $\lceil m \log_2 l \rceil$), and the technical assumption $l > 100$ can be made without loss of generality. The condition $s = 0.08t - q$ for $q := 1.5mn + m$ means that Eve can store about 8% of the randomizer, i.e. $\nu = 0.08$. Note that q is roughly equal to the number of randomizer bits the honest players need to access, and hence can be neglected. The right-hand side of (6) is negligible in m as long as $\log_2 n \leq cm$ (for any constant $c < \frac{1}{2}$), which is a very weak assumption.[8]

Proof (of Corollary 9). Let $\xi = \tau = 1/2$ in Theorem 8. It follows directly from (4) and $m \log_2 l \leq n$ that

$$\begin{aligned} \Delta &= lm(\log_2 e)/16 - m(1 + 0.5\,(m \log_2 l + n + 1)) \\ &\geq 0.09lm - mn - 1.5m \end{aligned}$$

Thus $s - \Delta \leq -0.01lm - 0.42mn - 0.08m \leq -0.01lm$ and hence, using $H_\infty(R) = t$, Theorem 8 yields

$$d(X|UK) \leq n(2^{-0.01lm} + 2^{-m/2}/2).$$

For $l > 100$ we have $2^{-0.01lm} < 2^{-m} < 2^{-m/2}/2$ and the corollary follows. □

Example 10 *This example was discussed in Section 2.1. For $s = 10^{15}, m = 125, n = 10^{10}$, and $l = 10^{14}$, we have $t \approx 1.25 \cdot 10^{16}$ and the length of the initial key has approximately 6000 bits. We obtain $d(X|UK) < 2^{-29}$.*

4.3 Proof of the Main Theorem

In this section we prove Theorem 8. At a high level, the proof consists of three steps. First, observe that from Lemma 3, to bound the value of $d(X)$ it suffices to prove a bound on the advantage of the optimal strategy for guessing X_η, given $X_1 \cdots X_{\eta-1}$, U, and K, for any $\eta = 1, \ldots, n$ and any storage function h. Second, for any fixed h and η and any fixed guessing strategy (called g) one can consider, for each value r of the randomizer, the average (over values of K) advantage of g in guessing X_η. Lemma 11 shows that the fraction of values r for which this average guessing advantage exceeds ϵ (where ϵ is a very small value) is negligible. Third, this fact implies by a standard argument that the overall guessing advantage is negligible.

[8] This assumption is actually optimal up to a constant. More precisely, for any $\nu \leq 1$ the scheme is insecure if $n \geq 1/\nu^m$ since, if Eve would store a fraction ν of the bits of each block, the probability that she could compute at least one bit of the expanded key would be substantial.

The overall structure of the proof has similarities with Trevisan's [24] work (see also Shaltiel's overview article [22]). The basic argument there is also to show that the number of inputs (to the extractor) for which there exists a good next-bit predictor for the extractor output, is negligible. However, our proof of the second step (i.e., the proof of Lemma 11) seems to use fundamentally different techniques than those used in [24, 22].

For every h we have, according to Lemma 1,

$$\tfrac{1}{2} + d(X_\eta | X_1 \cdots X_{\eta-1} U K) = \max_{g\colon \{0,1\}^{\eta-1} \times \mathcal{U} \times \mathcal{K} \to \{0,1\}} P\left(g(X_1, \ldots, X_{\eta-1}, U, K) = X_\eta\right), \quad (7)$$

where g is a function guessing X_η when given $X_1, \ldots, X_{\eta-1}, U$, and K. We can now state the main technical lemma of the paper, which is proved in Section 5.

Lemma 11 *For every $\eta \in \{1, \ldots, n\}$, for every function $g : \{0,1\}^{\eta-1} \times \mathcal{U} \times \mathcal{K} \to \{0,1\}$, and for all $u \in \mathcal{U}$, $\tau \in [0,1]$, and $\xi \in [0,1]$, the fraction of randomizers $r \in \mathcal{R}$ for which*

$$P\left(g(X_1, \ldots, X_{\eta-1}, u, K) = X_\eta \,\Big|\, R = r\right) \geq \tfrac{1}{2} + \epsilon \quad (8)$$

is at most $2^{-\Delta}$, with Δ and ϵ defined by (4) and (5), respectively.

Let us briefly discuss Lemma 11. Consider a fixed η. The adversary's strategy (i.e., the choice of h and g) can be considered fixed, i.e., before R and K are chosen. The adversary is now assumed to be given $X_1, \ldots, X_{\eta-1}$ and K, and the purpose of g is to guess X_η when given $X_1, \ldots, X_{\eta-1}, K$, as well as the stored value $U = h(R)$. The concrete value u is here considered as an input to g, but we can als interpret g as a set of functions $g_u : \{0,1\}^{\eta-1} \times \mathcal{K} \to \{0,1\}$ with parameter $u \in \mathcal{U}$.

We analyse and bound (to close to 1) the fraction of randomizers r for which the (any) fixed choice of h and g is bad for the adversary, where bad means that the advantage in guessing X_η correctly is less than ϵ. In other words, for a given value u stored by the adversary we analyse the fraction of r for which (8) does *not* hold. Equivalently, we can bound (to close to zero) the fraction, say σ, of randomizers r for which (8) is satisfied, for a given u, i.e., which are potentially good for the adversary. There are at most $|\mathcal{U}| \leq 2^s$ values of u, hence the fraction of randomizers r for which (8) is satisfied for some u is at most $2^s\sigma$ (which is at most $2^{-\Delta}$). Equivalently, $1 - 2^s\sigma$ is a lower bound on the fraction of randomizers which are bad for the adversary for *all* $u \in \mathcal{U}$. For such an r, the event that $R = r$ is bad for the adversary even if an oracle would tell him the best possible value u he could have stored.

More precisely, by counting only those r that are bad for the adversary for *all* u, we need not consider the particular storage function h. Rather, we can allow the adversary (only for the purpose of the analysis) a more generous strategy which still does not help him. In this view, the adversary can choose $|\mathcal{U}| \leq 2^s$ different guessing functions g_u and is considered to be successful (for r) if any one (without him being required to know which one) of the g_u has an advantage $P\left(g_u(X_1, \ldots, X_{\eta-1}, K) = X_\eta \mid R = r\right) \geq \frac{1}{2} + \epsilon$.

To continue the proof of Theorem 8, let Γ be the set of values $r \in \mathcal{R}$ for which (8) is satisfied for some $u \in \mathcal{U}$. We have

$$|\Gamma| \leq |\mathcal{U}| \cdot 2^{-\Delta} |\mathcal{R}| \leq 2^{s+t-\Delta}.$$

Since the maximal probability of any $r \in \mathcal{R}$ is $2^{-H_\infty(R)}$, this yields

$$P(R \in \Gamma) \leq 2^{s+t-H_\infty(R)-\Delta}.$$

For all $r \notin \Gamma$ we have $P(g(X_1, \ldots, X_{\eta-1}, u, K) = X_\eta | R = r) < \frac{1}{2} + \epsilon$. Therefore

$$P(g(X_1, \ldots, X_{\eta-1}, u, K) = X_\eta | R \notin \Gamma) < \tfrac{1}{2} + \epsilon.$$

We hence obtain, for all g,

$$\begin{aligned} P\left(g(X_1, \ldots, X_{\eta-1}, U, K) = X_\eta\right) &\leq \tfrac{1}{2} + \epsilon + P(R \in \Gamma) \\ &\leq \tfrac{1}{2} + 2^{s+t-H_\infty(R)-\Delta} + \epsilon. \end{aligned}$$

Thus, using (7) we get

$$d(X_\eta | X_1 \cdots X_{\eta-1} U K) \leq 2^{s+t-H_\infty(R)-\Delta} + \epsilon. \tag{9}$$

Since the choice of η was arbitrary, (9) holds for any η. Thus we get

$$\begin{aligned} d(X|UK) &\leq \sum_{\eta=1}^{n} d(X_\eta | X_1 \cdots X_{\eta-1} U K) \\ &\leq n \cdot \left(2^{s+t-H_\infty(R)-\Delta} + \epsilon\right), \end{aligned}$$

where the first step follows from Lemma 3. This completes the proof of Theorem 8. □

5 Proof of the Main Technical Lemma

To prove Lemma 11, which states an upper bound on the number of randomizers $r \in \mathcal{R}$ with property (8), we consider throughout this section (and also in Section 6) a new random experiment E in which R is chosen *uniformly* from $\mathcal{R}$, and K is chosen as before. This allows to convert a statement about the probability that R has property (8) (in this new random experiment) directly into a statement about the number of $r \in \mathcal{R}$ satisfying (8) in the original (or the new) random experiment, which we are interested in.

Let us now fix, throughout the section, an arbitrary guessing function $g : \{0,1\}^{\eta-1} \times \mathcal{U} \times \mathcal{K} \to \{0,1\}$ and an arbitrary value $u \in \mathcal{U}$. Let $c : \mathcal{K} \times \mathcal{R} \to \{-1, 1\}$ be the function that indicates for given key $k \in \mathcal{K}$ and randomizer $r \in \mathcal{R}$ whether or not g guesses the nth bit of the expanded key correctly when given the first $\eta - 1$ bits (as well as u and k):

$$c(k, r) := \begin{cases} 1 & \text{if } g(f_1(r, k), \ldots, f_{\eta-1}(r, k), u, k) = f_\eta(r, k) \\ -1 & \text{otherwise,} \end{cases}$$

where $f_j : \mathcal{R} \times \mathcal{K} \to \{0, 1\}$ is the function computing the jth bit of the expanded key (i.e., $f(r, k) = (f_1(r, k), \ldots, f_n(r, k))$ and $X_j = f_j(R, K)$).

We give a high-level overview of the main ideas of the proof. For each fixed $r \in \mathcal{R}$ we are interested in

$$P\left(g(X_1, \ldots, X_{\eta-1}, u, K) = X_\eta \;\middle|\; R = r\right) = \tfrac{1}{2} + \tfrac{1}{2} \underbrace{l^{-m} \sum_{k \in \mathcal{K}} c(k, r)}_{=:\alpha(r)}.$$

Let $\alpha(r) := l^{-m} \sum_{k \in \mathcal{K}} c(k, r)$. Lemma 11 is thus equivalent to

$$P\left(\alpha(R) \geq 2\epsilon\right) \leq 2^{-\Delta}. \tag{10}$$

The quantity $\alpha(r)$ and the following analysis is best understood by considering the complete l-ary tree of depth m, with the l^m leaves labeled by the keys $k \in \mathcal{K}$. Each leaf labeled k is assigned the value $c(k,r)$, and $\alpha(r)$ is simply the average of these values assigned to the leaves.

Each node at depth i in the tree can be labeled by a key prefix of length i, i.e., by an element $\kappa \in \mathcal{K}_i$, where $\mathcal{K}_i := \{1,\ldots,l\}^i$. To each node κ in the tree we can assign as its value the average of the leaf values in the corresponding subtree, denoted $\alpha_\kappa(r)$. It can be computed by adding the values of the l sons of node κ and dividing the sum by l:

$$\alpha_\kappa(r) \;:=\; l^{|\kappa|-m} \sum_{k\in\mathcal{K} \text{ with prefix } \kappa} c(k,r) \;=\; \frac{1}{l}\sum_{a=1}^{l} \alpha_{\kappa\|a}(r), \tag{11}$$

where $\|$ denotes concatenation. Note that

$$-1 \le \alpha_\kappa(r) \le 1 \tag{12}$$

and typically (as we want to prove), $\alpha_\kappa(r)$ is very close to 0. The root value thus obtained is $\alpha(r)$ (where we can omit the empty string as subscript).

To make the analysis of the distribution of the random variable $\alpha(R)$ (required to prove (10)) manageable we need to make simplifications, which of course must be conservative in the sense that they cannot decrease $\alpha(R)$. For this purpose we define $\beta_i(r)$ (for $0 \le i \le m$) as the maximum of all absolute values of $\alpha_\kappa(r)$, maximized over $\kappa \in \mathcal{K}_i$ and over choices of the first i rows of r:

$$\beta_i(r) := \max_{\kappa\in\mathcal{K}_i}\max_{\rho\in\mathcal{R}_i} |\alpha_\kappa((\rho, r(i+1),\ldots,r(m)))|\,, \tag{13}$$

where $\mathcal{R}_i := \{0,1\}^{i\times(l+n-1)}$ denotes the set of binary $i\times(l+n-1)$-matrices, $r(i)$ denotes the ith row of randomizer r, and $(\rho, r(i+1),\ldots,r(m)) \in \mathcal{R}$ denotes the randomizer consisting of ρ as the first i rows and $r(i+1),\ldots,r(m)$ as the lower $m-i$ rows. Note that $\beta_i(r)$ is actually a function only of the lower $m-i$ rows of r and is independent of the first i rows (because of the maximization over ρ).

We define A_i (for $0 \le i \le m$) as the random variable

$$A_i := \beta_i(R). \tag{14}$$

Since $\beta_0(r) = \alpha(r)$ we have $A_0 = \alpha(R)$. Also, $\beta_m(r) = 1$ for all r and hence $A_m = 1$. Because of (11) and (12) we have

$$\beta_i(r) \le \beta_{i+1}(r)$$

for all r and thus

$$A_i \le A_{i+1}. \tag{15}$$

Thus the reverse sequence $A_m,\ldots,A_0$ of random variables is non-increasing, and our goal is to show that, in fact, in each step it decreases significantly (i.e., is multiplied by $\tau \ll 1$) with high probability (i.e., with probability $1-\pi$ for a small π).

Note that $A_0,\ldots,A_m$ are defined in our random experiment of selecting $R = (R(1),\ldots,R(m))$ and K, but the random variable A_i actually depends only on $R(i+1),\ldots,R(m)$. For example, A_m is constant ($=1$).

For a key prefix $\kappa \in \mathcal{K}_i$ of some length i, let $w(\kappa,r)$ be the n-bit string obtained as the XOR of the i n-bit sub-key blocks selected by κ in the first i rows of r. In other words, $w(\kappa,r)$

is the expanded key that would result if a scheme restricted to the first i rows would be used. It will be crucial later on that for fixed κ, $\alpha_\kappa(r)$ and hence $\beta_i(r)$ depends on ρ (the first i rows of r) only through $w(\kappa, r)$ (actually only through the first η bits of $w(\kappa, r)$). Therefore, for a given κ the maximization over ρ in (13) can be seen as a maximization only over the 2^n values of $w(\kappa, (\rho, \overline{\rho}))$ (for any $\overline{\rho}$).[9]

Lemma 12 *The random variables $A_0, \ldots, A_m$ satisfy, for every $\tau \in [0,1]$ and for $0 \leq i \leq m-1$:*

$$P\left(A_i \geq \tau A_{i+1} \,\middle|\, A_{i+1} = a_{i+1}, \ldots, A_m = a_m\right) \leq \pi \tag{16}$$

for all $a_{i+1}, \ldots, a_m$,[10] *where*

$$\pi := l^m 2^{n+1} e^{-l\tau^2/2}. \tag{17}$$

Let us discuss the implications of this lemma before proving it. It implies that with very high probability the value of $A_0 = \alpha(R)$ is very small. More precisely, we have the following lemma.

Lemma 13 *If, for some $\pi, \tau \in [0,1]$, a sequence of positive-valued real random variables $A_0, \ldots, A_m$ with $A_m = 1$ satisfies (15) and (16), then, for every $\xi \in [0,1]$ we have*

$$P\left(A_0 \geq \tau^{\xi m}\right) \leq 2^m \pi^{(1-\xi)m}.$$

Proof. For $0 \leq i \leq m-1$, let

$$B_i := \begin{cases} 1 & \text{if } A_i \leq \tau A_{i+1} \\ 0 & \text{if } A_i > \tau A_{i+1}. \end{cases}$$

Because of (15) we have $A_0 \leq \tau^{B_1 + \cdots + B_m}$ and hence it suffices to prove

$$P\left(B_1 + \cdots + B_m \leq \xi m\right) \leq 2^m \pi^{(1-\xi)m}. \tag{18}$$

Inequality (16) implies that for every i and for all $b_{i+1}, \ldots, b_m \in \{0,1\}$,

$$P\left(B_i = 0 \,\middle|\, B_{i+1} = b_{i+1}, \ldots, B_m = b_m\right) \leq \pi$$

and hence, for all $b_1, \ldots, b_m \in \{0,1\}$,

$$P\left(B_1 = b_1, \ldots, B_m = b_m\right) \leq \pi^{|\{i:\, b_i = 0\}|} = \pi^{m - (b_1 + \cdots + b_m)}. \tag{19}$$

Thus the event $B_1 + \cdots + B_m \leq \xi m$ of (18) is the union of the events $(B_1 = b_1 \wedge \cdots \wedge B_m = b_m)$ for those $b_1, \ldots, b_m$ with $b_1 + \cdots + b_m \leq \xi m$. Using (19), each of these events has probability at most $\pi^{m - \xi m}$, and there are trivially at most 2^m such events. Hence (18) follows and Lemma 13 is proved.[11] □

[9] Actually, one could consider only the 2^η values of the first η bits of $w(\kappa, (\rho, \overline{\rho}))$, but we do make use of the resulting slightly stronger bound.

[10] Formally, one should require the conditioning event to have non-zero probability, but here and in the sequel we will for simplicity omit such conditions.

[11] One could also bound $P\left(B_1 + \cdots + B_m \leq \xi m\right)$ by the probability $P\left(C_1 + \cdots + C_m \leq \xi m\right)$ where $C_1, \ldots, C_m$ are independent binary random variables with $P(C_i = 0) = \pi$. However, the improvement in the bound would only be minor.

Proof (of Lemma 12). Consider a fixed $i \in \{0, \ldots, m-1\}$. We prove that (16) holds even when further conditioned on the event, denoted $\mathcal{E}_{\rho'}$, that $(R(i+2), \ldots, R(m)) = \rho'$, for any fixed value $\rho' \in \mathcal{R}_{m-i-1}$ of the bottom $m-i-1$ blocks of the randomizer. In other words, we prove $P(A_i \geq \tau A_{i+1} | \mathcal{E}_{\rho'}) \leq \pi$ for all ρ'. Hence (16) also holds (as stated), i.e., without further condition. Note that $A_{i+1}, \ldots, A_m$ are determined by ρ'; let these values be $a_{i+1}, \ldots, a_m$, respectively. Therefore to prove (16) we need to show that for all ρ'

$$P\left(\max_{\kappa \in \mathcal{K}_i} \max_{\rho \in \mathcal{R}_i} |\alpha_\kappa((\rho, R(i+1), \rho'))| \geq \tau a_{i+1}\right) \leq \pi. \tag{20}$$

This probability can be computed in a restricted random experiment defined only by the row $R(i+1)$ since the rest is fixed. Let us fix some particular values of $\kappa \in \mathcal{K}_i$, $\rho \in \mathcal{R}_i$ and $\rho' \in \mathcal{R}_{m-i-1}$, and investigate the event that

$$|\alpha_\kappa((\rho, R(i+1), \rho'))| \leq \tau a_{i+1}.$$

Applying (11) yields

$$\alpha_\kappa((\rho, R(i+1), \rho')) = \frac{1}{l} \sum_{j=1}^{l} S_j, \tag{21}$$

where

$$S_j := \alpha_{\kappa \| j}((\rho, R(i+1), \rho')).$$

The random variables S_j are not independent, but they are sufficiently independent in the sense of the following lemma, which we need to continue the proof of Lemma 12.

Lemma 14 *$S_1, \ldots, S_l$ is a martingale difference sequence.*

Proof. First, observe that for every j the bit $R(i+1, j+\eta-1)$ is independent of $(S_1, \ldots, S_{j-1})$. Flipping the bit $R(i+1, j+\eta-1)$ complements the η-th bit of the expanded key for any initial key k with prefix $\kappa \| j$. Hence this also flips all the values of $c(k, (\rho, R(i+1), \rho'))$ for such k and thus changes the sign of $\alpha_{\kappa \| j}((\rho, R(i+1), \rho'))$, i.e.,

$$\alpha_{\kappa \| j}((\rho, R(i+1), \rho')) = -\alpha_{\kappa \| j}((\rho, \hat{R}(i+1), \rho'))$$

where $\hat{R}(i+1)$ is obtained from $R(i+1)$ by flipping bit $R(i+1, j+\eta-1)$. Therefore the probability distribution $P_{S_j | S_1 \cdots S_{j-1}}$ is symmetric in the sense that

$$P_{S_j | S_1 \cdots S_{j-1}}(s_j, s_1, \ldots, s_{j-1}) = P_{S_j | S_1 \cdots S_{j-1}}(-s_j, s_1, \ldots, s_{j-1}) \tag{22}$$

for all $s_1, \ldots, s_j \in \{0, 1\}$. This implies that $E[S_j \mid S_1 = s_1, \ldots, S_{j-1} = s_{j-1}] = 0$ for all $s_1, \ldots, s_{j-1}$, i.e., $S_1, \ldots, S_l$ is a martingale difference sequence. □

Let us return to the proof of Lemma 12. From Lemma 6 and the fact (because of (11)) that $|S_j| \leq a_{i+1}$ for all j we obtain

$$P\left(\left|\sum_{j=1}^{l} S_j\right| \geq \tau l a_{i+1}\right) \leq 2e^{-l\tau^2/2}, \tag{23}$$

which together with (21) implies

$$P\left(|\alpha_\kappa((\rho, R(i+1), \rho'))| \geq \tau a_{i+1}\right) \leq 2e^{-l\tau^2/2}. \tag{24}$$

Note that (24) holds for any $\kappa \in \mathcal{K}_i$, $\rho \in \mathcal{R}_i$ and $\rho' \in \mathcal{R}_{m-i-1}$, as our choice of these values was arbitrary. To finalize the proof of Lemma 12 we observe that, as explained above, the maximization in (20) is only over the l^i values of κ and over the 2^n values of $w(\kappa, (\rho, \overline{\rho}))$ (where the value of $\overline{\rho}$ is irrelevant). The event

$$\max_{\kappa \in \mathcal{K}_i} \max_{\rho \in \mathcal{R}_i} \left|\alpha_\kappa((\rho, R(i+1), \rho'))\right| \le a_{i+1}$$

is thus the union of $l^i \cdot 2^n$ events of the form $|\alpha_\kappa((\rho, R(i+1), \rho'))| \le a_{i+1}$, each of which has probability at most $2e^{-l\tau^2/2}$, according to (24). Lemma 12 now follows because $i < m$. $\square$

Proof (of Lemma 11). Since $\alpha(R) = A_0$, Lemma 13 together with (17) implies

$$P\left(\alpha(R) \ge \tau^{\xi m}\right) \le 2^{m+(1-\xi)m(m\log_2 l+n+1)} e^{-\tau^2(1-\xi)lm/2},$$

which implies (10) for Δ given by (4) by changing the base in the right term from e to 2. This completes the proof of Lemma 11. $\square$

6 Asymptotically Optimal Randomness Efficiency

In this section we prove that the security of our scheme can be proved for randomness efficiency arbitrarily close to 1. We assume that the min-entropy of R is at least some constant fraction γ of the maximal value t, i.e.,

$$H_\infty(R) \ge \gamma t.$$

The randomness efficiency ν (which can generally depend on γ) can be defined as

$$\nu := \frac{s}{H_\infty(R)}.$$

Hence $s \le \gamma\nu t$. Let

$$\psi(m, l, n, \gamma, \nu) := \max_{P_R:\, H_\infty(R) \ge \gamma t} \; \max_{h:\, \{0,1\}^t \to \{0,1\}^{\lfloor \gamma\nu t \rfloor}} d(X|UK)$$

be the maximal average distance of X from uniform that the adversary can achieve. To state an asymptotic result we must fix how the parameters l, m and n grow with respect to each other. Let l and n be functions of m. First, we let $l := \lambda(m)$ where λ is any function satisfying

$$m^3 \le \lambda(m) \le 2^m. \tag{25}$$

(The choice of these bounds is rather arbitrary.) Second, we consider any fixed key expansion factor c, i.e., $n := \lfloor cm \log_2 l \rfloor$. From (25) we have $n \le cm^2$ and

$$t = m(l + n - 1) = m\lambda(m) + o(m\lambda(m)).$$

Theorem 15 *For every $\gamma \in (0, 1], \nu \in [0, 1), c > 1$ and λ satisfying (25),*

$$\psi(m, \lambda(m), cm \log_2(\lambda(m)), \gamma, \nu) = 2^{-\Omega(m)}.$$

The theorem, stating that $\psi(m, \lambda(m), cm\log_2(\lambda(m)), \gamma, \nu)$ decreases exponentially in m, is primarily of theoretical interest since for ν close to 1 the parameter m must be quite large. Let us briefly explain why Theorem 8 does not imply Theorem 15. It is not difficult to see that for $\gamma = 1$ (i.e., $H_\infty(R) = t$) and for all ξ and τ we have $\Delta \leq lm(\log_2 e)/2$. For $s \geq \Delta$ the right hand side of (3) is at least n. Thus Theorem 8 gives no non-trivial bound on $d(X|UK)$ if ν exceeds $lm(\log_2 e)/2t \approx (\log_2 e)/2 \approx 0.72$. The factor $(\log_2 e)/2$ comes from Lemma 6 used to prove (23). We therefore need Lemma 16 as an alternative to Lemma 6.

Lemma 16 *Let $S_1, \ldots, S_l$ be a sequence of random variables such that for $1 \leq j \leq l$ we have $|S_j| \leq a$ (for some $a \geq 0$) and (22) is satisfied, then, for every $\tau \in [0.5, 1)$,*

$$P\left(\left|\sum_{j=1}^{l} S_j\right| \geq \tau l a\right) \leq 2^{-l(1-H(\tau))+1}, \tag{26}$$

where $H(\tau) := -\tau\log_2\tau - (1-\tau)\log_2(1-\tau)$ is the binary entropy function.

This lemma is incomparable with Lemma 6. On one hand it is weaker since it applies only when (22) is satisfied and for $\tau \geq \frac{1}{2}$. Also, for τ not much greater than 1/2 (26) is worse than (23). However, for our purposes (τ close to 1), (26) is stronger than (23).

Proof (of Lemma 16). For every j let $T_j := |S_j|$ and

$$V_j = \begin{cases} 1 & \text{if } S_j \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

Thus $S_j = T_j \cdot V_j$. Therefore, for every τ, $\sum_{j=1}^{l} S_j \geq \tau l a$ implies $|\{j : V_j = 1\}| \geq \tau l$. The random variables $V_1, \ldots, V_l$ are independent and uniformly distributed and thus $|\{j : V_j = 1\}|$ is distributed according to the binomial distribution. Therefore

$$\begin{aligned} P\left(\sum_{j=1}^{l} S_j \geq \tau l a\right) &\leq P\left(|\{j : V_j = 1\}| \geq \tau l\right) \\ &\leq 2^{-l}\sum_{j=0}^{(1-\tau)l}\binom{l}{j} \\ &\leq 2^{-l(1-H(1-\tau))}. \end{aligned} \tag{27}$$

The last step follows from $\sum_{j=0}^{\omega l}\binom{l}{j} \leq 2^{lH(\omega)}$ (see for instance [26], Theorem 1.4.5, page 21) which holds for all $\omega \in (0, 0.5]$, hence for $\omega = 1 - \tau$. By symmetry we also get the same bound as in (27) for $P\left(\sum_{j=1}^{l} S_j \leq -\tau l a\right)$ and thus, by multiplying the bound by 2 and using $H(1-\tau) = H(\tau)$, (26) follows. □

If we now use Lemma 16 instead of Lemma 6 in the proof of Theorem 8 we can replace the term $\tau^2(\log_2 e)/2$ in (4) by $1 - H(\tau)$ and thus obtain the following lemma which is incomparable to Theorem 8.

Lemma 17 *Theorem 8 also holds for $\tau \geq \frac{1}{2}$ if (4) is replaced by*

$$\Delta := lm(1-\xi)(1-H(\tau)) - m(1 + (1-\xi)(m\log_2 l + n + 1)).$$

We are now ready to prove Theorem 15.

Proof (of Theorem 15). We have $s + t - H_\infty(R) \leq (\gamma\nu + 1 - \gamma)t = \zeta t$ for $\zeta := 1 - \gamma(1-\nu)$. Lemma 17 and $n \leq cm^2$ imply that for every $\tau \geq \frac{1}{2}$ and ξ,

$$\psi(m, \lambda(m), cm\log_2(\lambda(m)), \gamma, \nu) \leq cm^2 \left(2^{\zeta t - \Delta} + \epsilon\right).$$

Note that

$$\zeta t - \Delta = (\zeta - (1-\xi)(1 - H(\tau)))m\lambda(m) + o(m\lambda(m)).$$

By choosing $\tau \geq \frac{1}{2}$ and ξ such that $\zeta - (1-\xi)(1-H(\tau)) < 0$ it follows that $2^{\zeta t - \Delta}$ vanishes (more than) exponentially in m, as does ϵ and hence also the right-hand side of the above inequality. □

7 Conclusions and Open Problems

The problem of finding very simple key-expansion schemes for the bounded-storage model, with very short initial keys and optimal randomizer efficiency (also concrete, non-asymptotic), remains an interesting research topic. This question, in particular the notion of simplicity, is necessarily related to the future developments in the communication and storage technologies.

A very interesting question is whether our scheme or any of the other proposed schemes remains secure against an adversary who can store s quantum bits rather than s classical bits of information. This question is relevant because the read-out operation on the state of the quantum memory could depend on the initial key K which the adversary is assumed to learn after having performed the store operation, and this makes quantum memory potentially more powerful than classical memory. This issue is discussed in [15], where it is proved for instance that privacy amplification by universal hashing [5], previously proved secure against storage-bounded but otherwise unbounded classical adversaries, remains equally secure even against an adversary with quantum memory. These results do not seem to carry over to the bounded-storage setting, but we conjecture that security in the bounded-storage model holds also against quantum adversaries.

Acknowledgments

We would like to thank Oded Goldreich and the anonymous referees for many helpful suggestions and for drawing our attention to [24] and [22], and Rasmus Pagh for a helpful discussion.

References

1. Y. Aumann, Y. Z. Ding, and M. O. Rabin. Everlasting security in the bounded storage model, *IEEE Transactions in Information Theory*, 48:1668–1680, 2002.
2. Y. Aumann and M. O. Rabin. Information theoretically secure communication in the limited storage space model. In *Advances in Cryptology - CRYPTO '99*, volume 1666 of *Lecture Notes in Computer Science*, pp. 65–79. Springer-Verlag, 1999.
3. K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Math. Journal*, 19:357–367, 1967.
4. C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5(1):2–28, 1992.
5. C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, 1995.

6. C. Cachin, C. Crépeau, and J. Marcil. Oblivious transfer with a memory-bounded receiver. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 493–502. IEEE Computer Society, 1998.
7. C. Cachin and U. Maurer. Unconditional security against memory-bounded adversaries. In Burton S. Kaliski Jr., editor, *Advances in Cryptology - CRYPTO '97*, volume 1294 of *Lecture Notes in Computer Science*, pp. 292–306. Springer-Verlag, 1997.
8. B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.
9. T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley Series in Telecommunications. Wiley International, 1992.
10. Y. Z. Ding. Oblivious transfer in the bounded storage model. In *Advances in Cryptology - CRYPTO 2001*, volume 2139 of *Lecture Notes in Computer Science*, pp. 155–170. Springer-Verlag, 2001.
11. Y. Z. Ding. *Provably Everlasting Security in the Bounded Storage Model.* Ph.D. Thesis, Harvard University, 2001.
12. Y. Z. Ding and M. O. Rabin. Hyper-encryption and everlasting security. In *19th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 2285 of *Lecture Notes in Computer Science*, pp. 1–26. Springer-Verlag, 2002.
13. S. Dziembowski and U. Maurer. Tight security proofs for the bounded-storage model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 341–350. ACM, May 2002.
14. S. Goldwasser and S. Micali. Probabilistic Encryption. *Journal of Computer and System Sciences*, 28:270–299, 1984.
15. R. König, U. Maurer, and R. Renner. On the power of quantum memory. Preprint, Quantum physics e-print archive, `http://arxiv.org/abs/quant-ph/0305154`, 2003.
16. C.-J. Lu. Encryption against storage-bounded adversaries from on-line strong extractors. To appear in *Journal of Cryptology* (this issue).
17. J.L. Massey and I. Ingemarsson. The Rip van Winkle cipher - a simple and provably computationally secure cipher with a finite key. In *Proc. IEEE Int. Symp. Information Theory (Abstracts)*, page 146, 1985.
18. U. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *Journal of Cryptology*, 5(1):53–66, 1992. (Conference version appeared in Proc. of Eurocrypt '90.)
19. U. Maurer. Secret key agreement by public discussion. *IEEE Transaction on Information Theory*, 39:733–742, 1993.
20. R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.
21. N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43-52, 1996.
22. R. Shaltiel. Recent Developments in Extractors. *Bulletin of the European Association for Theoretical Computer Science*, 77:67-95, 2002.
23. C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.
24. L. Trevisan. Constructions of extractors using pseudorandom generators. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, pp. 141–148. ACM, May 1999.
25. S. P. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded storage model. To appear in *Journal of Cryptology* (this issue). See also Cryptology ePrint Archive (`http://eprint.iacr.org/`, Report 2002/162, 2002.
26. J. H. van Lint. *Introduction to Coding Theory.* 3rd edition, Springer-Verlag, 1998.