Tight Security Proofs for the Bounded-Storage Model

Stefan Dziembowski[*] [†]
dziembowski@inf.ethz.ch

Ueli Maurer[‡]
maurer@inf.ethz.ch

Department of Computer Science
ETH Zurich, Switzerland

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, thus at first glance apparently contradicting Shannon's bound on the key size of a perfect cipher.

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 in fact to be longer than the derived one-time pad, or t had to be extremely large ($t > ns$), or the adversary was assumed to be able to store only actual bits of R rather than arbitrary s bits of information about R, or the adversary could obtain a non-negligible amount of information about X.

In this paper we prove the first non-restricted security result in the bounded-storage model, exploiting the full potential of the model: K is short, X is very long (e.g. gigabytes), t needs to be only moderately larger than s, and the security proof is optimally strong. In fact, we prove that s/t can be arbitrarily close to 1 and hence the storage bound is essentially optimal.

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.

This paper is concerned with provably-secure key-agreement schemes *not* based on an intractability assumption. 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 therefore concentrate on key agreement, i.e., on generating the one-time pad.

1.1 Assumptions in Cryptography and Information-Theoretic Security

The security of every cryptographic system depends on certain assumptions. Usually, not all assumptions are stated explicitly. For instance, two obvious such assumptions are (1) that randomness exists, i.e., that one can generate random keys, and (2) that such keys are independent of an adversary's view.[1]

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. Moreover, if the underlying model of computation is a classical computer (e.g. a Turing machine), this is not fully satisfactory in view of the new developments in quantum computing.

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,

[*] Part of this work was done while the first author was a Ph.D. student at BRICS, Århus University, Denmark.

[†] Supported in part by the Polish KBN grant no. 7 T11C 027 20.

[‡] Supported in part by the Swiss National Science Foundation, grant no. 2000-055466.98.

STOC'02, May 19-21, 2002, Montreal, Quebec, Canada.
Copyright 2002 ACM 1-58113-495-9/02/0005 ...$5.00.

[1] This implies, for instance, that it is impossible to read somebody's mind, at least not in a cryptographic context.

the security of such a system may rely on an assumption about the probabilistic behavior of nature, for instance of a noisy channel [15] or a quantum measurement [4].

A main goal of research in cryptography is to reduce the assumptions underlying a security proof.

1.2 Motivation of the Bounded-Storage Model

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 by in [14], is that one makes a sole (conservative) assumption about Eve's storage capacity (say 1000 Terabytes), but no assumption 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 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.[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 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 [10] or [9], Section 1.3.1 for an analysis of the current technology).

As a concrete example of what is proved in this paper (Example 1), 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 a key of length 10 Gigabits (i.e. $n = 10^{10}$) about which the adversary has essentially no information. Alice and Bob need to read only $1.25 \cdot 10^{12}$ bits from the random source.

1.3 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 decrypt (theoretically) by an exhaustive key search, provided the plaintext is redundant. 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.

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 can be solved by adding (bitwise modulo 2) several, say m, sub-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 [14], which we also use here and which was also 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 [14] each of the m sub-block is taken from a separate non-overlapping block of R, with a cyclic continuation if the sub-block reaches the end of the block, whereas in this paper no cyclic extension is used and in [2] the non-blocked scheme sketched above is used.

Massey and Ingemarsson proposed the so-called Rip van Winkle cipher [13] which is also a variation on the theme of book ciphers.

1.4 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 *derived key* $X = (X_1, \dots, X_n)$ (i.e. $n \gg \log_2 |R|$). This expansion at first glance apparently contradicts Shannon's and Maurer's bounds on the key size of a perfect or close to perfect cipher. Shannon [17] proved that if only one-way communication on the insecure channel (from the sender to the receiver) is allowed, the entropy of the secret key must be at least as large as the entropy of the encrypted message. It was proved in [15] that the bound holds also in the more natural scenario when Alice and Bob can communicate arbitrarily over an insecure channel.

In a first phase, a t-bit string R (chosen uniformly at random) 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. Alice and Bob apply a known *key-derivation function* $f : \mathcal{R} \times \mathcal{K} \to \{0,1\}^n$ to compute the derived 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[4] $h : \mathcal{R} \to \mathcal{U}$ for some $\mathcal{U}$ with the only restriction that $|\mathcal{U}| \leq 2^s$. The memory

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

[3] More generally, R must have more than s bits of entropy, but in this paper we do not consider non-uniform R.

[4] Without loss of generality we can consider only determin-

size during the evaluation of h need not be bounded. The value stored by Eve is $U = h(R)$. After storing U, Eve looses 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-derivation 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, one needs to prove that 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. Obviously, this cannot hold for $s \geq t$, but it should hold for as large a storage bound s as possible, ideally $s = \nu t$ for ν close to 1. The ratio

$$\nu := s/t$$

will be called the *randomness efficiency* of a scheme.

1.5 Previous Results for the Bounded-Storage Model

The bounded-storage model was introduced by Maurer in [14], but the proposed cipher was proved secure 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 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 [14], R is divided into m blocks of l bits, i.e. $t = lm$, and each bit of X is the XOR of m bits, one from each block. The key K determines which bits are XOR-ed. We refer to Section 3.1 for a precise description of the scheme analyzed in this paper, which is essentially the same as that of [14]. The main problem left open in [14] was to show 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 1.4).

Cachin and Maurer [6] 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 [14] was achieved by Aumann, Ding and Rabin [2, 1, 9, 10], using a scheme very similar to that of [14]. The core technical argument is a security proof for a scheme for generating a single key bit (i.e. $n = 1$) and for $\nu_{\mathrm{ADR}} \approx 0.2$. The proof uses elegant techniques based on an application of the Cauchy-Schwartz inequality. 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 derivation n times [2, 1, 9] 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 decreases inversely proportional to n.
- Execute the single-bit derivation n times [1, 9, 10] 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. At first, this appears to be a very theoretical result, but in fact it has practical significance, as (1) the security of the derived key is in some sense higher than the security of the initial key (because it is *everlasting* – see [1] for more on this) and (2) as shown in [9, 10] one initial key can be reused in several independent schemes (even if the adversary can adaptively learn the derived keys).

The bounded-storage model was also studied in the context of secure two-party computations (see [9, 8, 5]), message authentication and non-malleable encryption [9, 10].

1.6 Contributions of this Paper

The main open question of [14, 2, 1] is whether key-expansion 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 3 and 4) addresses parameter sizes that are closest to being of practical interest. Theorem 1 and Corollary 1 state that for reasonable parameter sizes, secure key-expansion is possible for $\nu \approx 0.1$. In the second part (Section 5) we prove, as a purely theoretical result, that ν can be arbitrarily close to 1.

2. PROBABILITY-THEORETIC PRELIMINARIES

Our main goal is to show that, from the adversary's viewpoint (i.e., given all her information) the derived key X is, with overwhelming probability, distributed according to an essentially uniform distribution. Closeness to uniformity is measured without loss of generality in terms of statistical distance.

Definition 1. *For a probability distribution p over an alphabet $\mathcal{U}$, the* statistical distance of p from the uniform distribution *is*

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

It is easy to see that if $d(P_X)$ is small, then X is close to uniform in terms of other uniformity measures, for instance the Shannon entropy (see e.g. [7]) is close to maximal. It is also not difficult to see that if Alice and Bob use X as

istic strategies of Eve.

[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. A somewhat unrealistic but illustrative example could be the surface of the moon whose irregularities are interpreted as a huge array of random bits.

[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.

a one-time pad, then the resulting encryption scheme is *semantically secure* (a notion introduced in [12], see also [9], Section 2.1). More precisely, for any two messages chosen by Eve, her advantage in distinguishing between the encryptions of the two messages is at most $d(P_X)$.

Consider a random variable $X = (X_1, \ldots, X_n)$ distributed according to a distribution P_X over the n-bit strings. Define

$$\mu_i(P_X) := \max_{g\,:\,\{0,1\}^{i-1}\to\{0,1\}} P\left(g(X_1, \ldots, X_{i-1}) = X_i\right) - \tfrac{1}{2}. \quad (1)$$

In other words $\mu_i(P_X) + \frac{1}{2}$ is the maximal probability of guessing X_i correctly when given $X_1, \ldots, X_{i-1}$. The following lemma is proved in Appendix B.

Lemma 1. $d(P_X) \leq \sum_{i=1}^{n} \mu_i(P_X)$.

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

3. THE MAIN THEOREM

3.1 Description of the Cipher

We propose a scheme in which the randomizer R consists of blocks $R(1), \ldots, R(m)$ (for some $m \geq 1$ called the *height* of the randomizer) of $l+n-1$ random bits each (for some $l \geq 1$ called the *width* of the randomizer). Hence $t = m(l+n-1)$ and $R \in \mathcal{R} = \{0,1\}^t$. The initial key $K = (K_1, \ldots, K_m) \in \mathcal{K} = \{1, \ldots, l\}^m$ selects one starting point within each block, and the derived key $X = (X_1, \ldots, X_n)$ is the componentwise XOR of the m sub-blocks of length n beginning at these starting points, i.e., $X = f(R, K)$ where for every $r \in \mathcal{R}$ and $k = (k_1, \ldots, k_m) \in \mathcal{K}$ we set

$$f(r,k) := (\oplus_{i=1}^{m} r(i,k_i), \ldots, \oplus_{i=1}^{m} r(i,k_i+n-1)) \quad (2)$$

(here $r(i,j)$ denotes the jth bit in the ith row of r). (See also Figure 1.) The random experiment consists of selecting $R \in \mathcal{R}$ and $K \in \mathcal{K}$ independently and uniformly at random. All probabilities considered in this paper are for events in this random experiment.

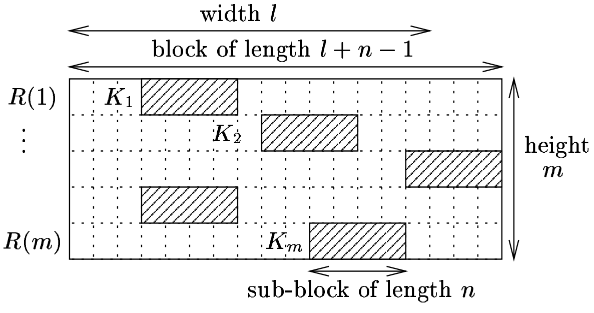

Figure 1: Illustration of the scheme for deriving a key $X = (X_1, \ldots, X_n)$, to be used as a one-time pad, from a short secret initial key $(K_1, \ldots, K_m)$. The randomizer is divided into m blocks $R(1), \ldots, R(m)$ of length $l+n-1$. The derived key X is the componentwise XOR of m sub-blocks length n, one selected from each block, where K_i is the starting point of the ith sub-block within the ith block $R(i)$.

3.2 Statement and Explanation of the Theorem

In this section we present and explain 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 $|\mathcal{U}| = 2^s$). Recall that

$$U = h(R)$$

is a random variable equal to the value stored in Eve's memory. Let

$$\begin{aligned} \beta(u,k) &:= d\left(P_{X|h(R)=u,K=k}\right) \quad (3)\\ &= d\left(P_{X|U=u,K=k}\right) \end{aligned}$$

be the bias of the distribution of the derived key X, given that $U = u$ and $K = k$. Hence $\beta(U,K)$ is a random variable whose value corresponds to the bias of Eve's distribution of the derived key X, given her entire information. (Remember that we assume that she can learn K after she looses the ability to access R.) The following theorem, the main result of the paper, states that the expected value of the bias is very small. The proof is given at the end of Section 3.3.

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

$$E\left[\beta(U,K)\right] \leq n \cdot (2^s \delta + \epsilon), \quad (4)$$

where

$$\delta := 2^{m+(1-\xi)m(m\log_2 l+n+1)-(\log_2 e)\tau^2(1-\xi)lm/2} \quad (5)$$

and

$$\epsilon := \tau^{\xi m}/2. \quad (6)$$

Expression (4) may be simplified if, as usual, we assume that $m, n \ll l$. In which case $\delta \approx e^{-\tau^2(1-\xi)lm/2}$.

Recall Markov's inequality (see e.g. [7], page 57) stating 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 1, namely that $E\left[\beta(U,K)\right]$ is negligible, into the statement that $\beta(u,k)$ can exceed a certain (very small) threshold only with negligible probability. The implication of Theorem 1 is stated below, as an example, for the concrete randomness efficiency $\nu = 0.08$.

Corollary 1. *If $l, m,$ and n are such that $m\log_2 l \leq n$ and $l > 100$, then, for $s := 0.08t - 1.5m(n+1)$ (where $t = m(l+n-1)$)),*

$$E\left[\beta(U,K)\right] \leq n2^{-m/2}. \quad (7)$$

The assumption $m\log_2 l \leq n$ reflects the fact that we are interested in key *expansion*[7], 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 essentially that Eve can store about 8% of the randomizer. Note that q is roughly equal to the number of randomizer bits the honest players need to access, and hence can be neglected. Observe also that the right-hand side of (7) is negligible in m as long

[7] The length of the initial key (written as a binary string) is $\lceil m\log_2 l\rceil$.

as $\log_2 n \leq cm$ (for any constant $c < \frac{1}{2}$), which is a very weak assumption.[8]

Proof of Corollary 1. Set $\xi = \tau = 1/2$ in Theorem 1. It follows directly from (5) (and using the assumption that $m \log_2 l \leq n$) that

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

Thus $s + \log_2 \delta \leq -0.01lm - 0.42mn - 0.08m \leq -0.01lm$ and hence Theorem 1 yields

$$E\left[\beta(U, K)\right] \leq n(2^{-0.01lm} + 2^{-m/2}/2).$$

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

Example 1. *This example was discussed in Section 1.2. Suppose $s = 10^{15}, m = 125, n = 10^{10}, l = 10^{14}$ (and thus $t \approx 1.25 \cdot 10^{16}$ and the length of the initial key is approximately 6000). Then $E\left[\beta(U, K)\right] < 2^{-29}$.*

3.3 Main Technical Lemma and the Proof of the Main Theorem

Consider the following guessing game. For some $j \in \{1, \ldots, n\}$ Eve must guess X_j when given $X_1, \ldots, X_{j-1}$ as well as U and K. Our goal is to prove that Eve's success probability for any j is at most negligibly greater than $1/2$. For a fixed storage function $h : \mathcal{R} \to \mathcal{U}$, Eve's strategy in this guessing game is characterized completely by a guessing function $g_j : \mathcal{U} \times \mathcal{K} \times \{0, 1\}^{j-1} \to \{0, 1\}$, where $g_j(U, K, X_1, \ldots, X_{j-1})$ is Eve's guess for X_j.[9] For every j, let

$$\alpha_j(u, k) := \mu_j(P_{X|U=u,K=k}).$$

From definition (1) of μ_j it follows that $\alpha_j(u, k) + \frac{1}{2}$ is equal to

$$\max_{g_j} \left(P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \;\middle|\; U = u, K = k \right) \right).$$

Hence

$$\begin{aligned} \max_{g_j} \left(P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \right) \right) & \\ = \tfrac{1}{2} + E\left[\alpha_j(U, K)\right]. & \end{aligned} \tag{8}$$

We can now state the main technical lemma of the paper. The proof is given in Section 4.

Lemma 2. *For all $j \in \{1, \ldots, n\}$, for every guessing function $g_j : \mathcal{U} \times \mathcal{K} \times \{0, 1\}^{j-1} \to \{0, 1\}$, for every $u \in \mathcal{U}$, $\tau \in [0, 1]$, and $\xi \in [0, 1]$, the fraction of randomizers r for which*

$$P\left(g_j(u, K, X_1, \ldots, X_{j-1}) = X_j \;\middle|\; R = r \right) \geq \tfrac{1}{2} + \epsilon \tag{9}$$

is at most δ, with δ and ϵ defined by (5) and (6), respectively.

Corollary 2. *For every $j \in \{1, \ldots, n\}$, and $\tau, \xi \in [0, 1]$ we have $E\left[\alpha_j(U, K)\right] \leq 2^s \delta + \epsilon$ (where ϵ and δ are defined by (5) and (6), respectively).*

Proof of Corollary 2. Fix any g_j. Let Γ be the set of all randomizers r for which there exists $u \in \mathcal{U}$ such that (9) holds. Clearly we have that

$$P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \right) \tag{10}$$

is equal to the sum of

$$P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \mid R \in \Gamma \right) \cdot P\left(R \in \Gamma \right) \tag{11}$$

and

$$P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \mid R \notin \Gamma \right) \cdot P\left(R \notin \Gamma \right). \tag{12}$$

From the union bound (over all $u \in \mathcal{U}$) applied to Lemma 2 we know that $P\left(R \in \Gamma \right) \leq |\mathcal{U}|\, \delta$. Thus (11) is at most $2^s \delta$. From the definition of Γ we have

$$P\left(g_j(U, K, X_1, \ldots, X_{j-1}) = X_j \;\middle|\; R \notin \Gamma \right) < \tfrac{1}{2} + \epsilon.$$

Thus (12) is smaller than $\frac{1}{2} + \epsilon$. Therefore (10) is at most $2^s \delta + \left(\frac{1}{2} + \epsilon \right)$. Applying (8) we get $E\left[\alpha_j(U, K)\right] \leq 2^s \delta + \epsilon$. □

Proof of Theorem 1. Lemma 1 implies that $\beta(u, k) \leq \sum_{j=1}^{n} \alpha_j(u, k)$. From the linearity of the expected value we get

$$E\left[\beta(U, K)\right] \leq \sum_{j=1}^{n} E\left[\alpha_j(U, K)\right],$$

which by Corollary 2 is at most $n \cdot (2^s \delta + \epsilon)$. □

4. PROOF OF THE MAIN TECHNICAL LEMMA

To prove Lemma 2 we introduce some more notation. For every randomizer r, initial key k and $j \in \{1, \ldots, n\}$, let $f_j(r, k)$ denote the jth bit x_j of the derived key $f(r, k)$. Let $\mathcal{R}_i$ denote the set of all randomizers of height i and width l (i.e., the set of binary $i \times (l + n - 1)$-matrices) and let $\mathcal{K}_i$ denote the set $\{1, \ldots, l\}^i$. For $\kappa \in \mathcal{K}_i$ and $a \in \{1, \ldots, l\}$ we have $\kappa \cdot a \in \mathcal{K}_{i+1}$, where the symbol '$\cdot$' denotes concatenation.

It suffices to consider the case $j = n$ because for our scheme guessing X_j when given $X_1, \ldots, X_{j-1}$ is as hard as guessing X_n when given $X_{n-j+1}, \ldots, X_{n-1}$, which is not easier than guessing X_n when given $X_1, \ldots, X_{n-1}$.

Let us now fix an arbitrary guessing function $g_n : \mathcal{U} \times \mathcal{K} \times \{0, 1\}^{n-1} \to \{0, 1\}$ and a value $u \in \mathcal{U}$. Define the function $c : \mathcal{K} \times \mathcal{R} \to \{-1, 1\}$ in the following way: for every $k \in \mathcal{K}$ and $r \in \mathcal{R}$ set

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

[8] This assumption is actually optimal up to a constant. More precisely, for any $z \geq 1$ the scheme is insecure if $n = z^m$ and the storage size s of Eve is equal to t/z. In this case Eve can simply store from each block $R(i)$ a set of $(l + n - 1)/z$ randomly chosen bits. Recall that for every j the jth bit of the generated key is equal to

$$X_j = \oplus_{i=1}^{m} R(i, K_i + j)$$

(cf. (2)). The chance that Eve stored all the bits $R(1, K_1 + j), \ldots, R(m, K_m + j)$ in her memory is z^{-m}. Therefore for every j the probability that Eve can compute X_j is z^{-m}. Thus the expected number of bits in $(X_1, \ldots, X_n)$ that Eve can compute is equal to $n \cdot z^{-m} = 1$.

[9] A reader familiar with the previous work in this field may observe that for a fixed u a function $g_1(u, \cdot) : \mathcal{K} \to \{0, 1\}$ can be viewed as an *answer vector* [2, 1].

In other words the value of $c(k,r)$ is equal to 1 if the function g guesses correctly the last bit of the generated key, for the randomizer r and the initial key k (and it is equal to -1 otherwise).

For every $r \in \mathcal{R}$ and $\kappa \in \mathcal{K}_i$ (where $i \leq m$) define the *advantage (of the guessing function g) with respect to a randomizer r and a key prefix κ* as:

$$\begin{aligned} \operatorname{adv}_\kappa(r) &:= E\left[c(K,r) \mid \kappa \text{ is a prefix of } K\right] \qquad (13) \\ &= l^{|\kappa|-m} \sum_{k \in \mathcal{K}\,:\,\kappa \text{ is a prefix of } k} c(k,r). \end{aligned}$$

When $|\kappa| = 0$ we will write $\operatorname{adv}(r)$ instead of $\operatorname{adv}_\kappa(r)$. For every randomizer r we have

$$\begin{aligned} P\Big(g_n(u, K, X_1, \ldots, X_{n-1}) = X_n \,\Big|\, R = r\Big) \\ = \tfrac{1}{2} + \tfrac{1}{2}\operatorname{adv}(r). \end{aligned}$$

Therefore, to prove Lemma 2, we need to show that

$$P\Big(\operatorname{adv}(R) \geq 2\epsilon\Big) \leq \delta. \qquad (14)$$

For every $i \in \{1, \ldots, m\}$, define the random variable A_i in the following way:

$$A_i := \max_{\kappa \in \mathcal{K}_i,\, \rho \in \mathcal{R}_i} \left|\operatorname{adv}_\kappa\left((\rho, R(i+1), \ldots, R(m))\right)\right|.$$ [10]

Note that A_i can be seen as defined in a random experiment determined only by the lowest $m-i$ blocks of the randomizer. But it is of course also defined in the random experiment of selecting the entire randomizer and the key at random, which we are interested in. Clearly $A_0 = \operatorname{adv}(R)$ and $A_m = 1$. Our goal is to show that A_i is always smaller than A_{i+1} (condition (15) below) and, moreover, that with very high probability A_i is smaller than τA_{i+1} for all values that $A_{i+1}, \ldots, A_m$ can take on (condition (16)).

LEMMA 3. *For every $\tau \in [0,1]$ and every i the sequence $A_0, \ldots, A_m$ satisfies the following conditions*

$$A_i \leq A_{i+1} \qquad (15)$$

and for every[11] $a_{i+1}, \ldots, a_m$

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

where

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

Before going to the proof let us present the implications of the lemma. It implies that with very high probability the value of A_0 is very small. More precisely, we have the following lemma.

LEMMA 4. *If (for some $\pi, \tau \in [0,1]$) a sequence of positive-valued real random variables $A_0, \ldots, A_m \in [0,\infty)$ (with $A_m = 1$) satisfies conditions (15), (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. Fix arbitrary values π, τ and ξ. For every $i \in \{0, \ldots, m-1\}$ set

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

Clearly (because of (15)) we have that

$$A_0 \leq \tau^{B_1 + \cdots + B_m}$$

Therefore it suffices to show that

$$P\Big(B_1 + \cdots + B_m \leq \xi m\Big) \leq 2^m \pi^{(1-\xi)m} \qquad (17)$$

From (16) we have that for every i and for all $b_{i+1}, \ldots, b_m$

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

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

$$\begin{aligned} P\left(B_1 = b_1, \ldots, B_m = b_m\right) &\leq \pi^{|\{i : b_i = 0\}|} \\ &= \pi^{m-(b_1+\cdots+b_m)}. \qquad (18) \end{aligned}$$

Set

$$\mathcal{B} := \left\{(b_1, \ldots, b_m) \in \{0,1\}^m : b_1 + \cdots + b_m \leq \xi m\right\}.$$

Thus

$$\begin{aligned} &P\left(B_0 + \cdots + B_m \leq \xi m\right) \\ &\quad = \sum_{(b_1,\ldots,b_m) \in \mathcal{B}} P\left(B_1 = b_1, \ldots, B_m = b_m\right) \\ &\quad \leq \sum_{(b_1,\ldots,b_m) \in \mathcal{B}} \pi^{m-\xi m} \qquad (19) \\ &\quad \leq 2^m \pi^{(1-\xi)m} \qquad (20) \end{aligned}$$

(where (19) follows from (18) and (20) follows from the trivial fact that $|\mathcal{B}| \leq 2^m$). Therefore (17) and and hence Lemma 4 is proved. □

We are now ready for the main technical argument.

PROOF OF LEMMA 3. Fix some $i \in \{0, \ldots, m-1\}$ and an arbitrary $\rho' = (\rho'_{i+2}, \ldots, \rho'_m) \in \mathcal{R}_{m-i-1}$. Consider the event that $(R(i+2), \ldots, R(m)) = \rho'$. Observe that the value $(a_{i+1}, \ldots, a_m)$ of $(A_{i+1}, \ldots, A_m)$ is determined by the fixed value ρ' of $(R(i+2), \ldots, R(m))$. More precisely for every $v \in \{i+1, \ldots, m\}$ we have

$$a_v = \max_{\bar\kappa \in \mathcal{K}_v,\, \bar\rho \in \mathcal{R}_v} \left|\operatorname{adv}_{\bar\kappa}\left((\bar\rho, \rho'_{v+1}, \ldots, \rho'_m)\right)\right|. \qquad (21)$$

Therefore to prove (16) we have to show that

$$P\left(\max_{\kappa \in \mathcal{K}_i, \rho \in \mathcal{R}_i} \left|\operatorname{adv}_\kappa\left((\rho, R(i+1), \rho')\right)\right| \geq \tau a_{i+1}\right) \leq \pi, \qquad (22)$$

and to prove (15) we have to show that

$$\max_{\kappa \in \mathcal{K}_i, \rho \in \mathcal{R}_i} \left|\operatorname{adv}_\kappa\left((\rho, R(i+1), \rho')\right)\right| \leq a_{i+1} \qquad (23)$$

[10] The expression $(\rho, R(i+1), \ldots, R(m))$ denotes the randomizer $r'' \in \mathcal{R}_m$ with $r''(j) := \rho(j)$ for $j \leq i$ and $r''(j) := R(j)$ for $j > i$. We will also write $(\rho, R(i+1), \rho')$ (for $\rho \in \mathcal{R}_i$ and $\rho' \in \mathcal{R}_{m-i-1}$) to denote the randomizer in $\mathcal{R}$ defined in a similar way.

[11] Formally, we should say: "for every $a_{i+1}, \ldots, a_m$ such that $P\left(A_{i+1} = a_{i+1}, \ldots, A_m = a_m\right) \geq 0$" (because otherwise (16) is not defined). For simplicity (here and in the sequel) we will not explicitly state conditions of this type.

(for all values of $R(i+1)$). Let us for a moment fix some particular values of $\kappa \in \mathcal{K}_i$ and $\rho \in \mathcal{R}_i$. Then from the definition (13) of the advantage we have

$$\begin{aligned}
&\mathrm{adv}_\kappa\left((\rho, R(i+1), \rho')\right) \\
&\quad= \frac{1}{l}\sum_{j=1}^{l} \mathrm{adv}_{\kappa\cdot j}\left((\rho, R(i+1), \rho')\right) \\
&\quad= \frac{1}{l}\sum_{j=1}^{l} S_j, \qquad (24)
\end{aligned}$$

where (to simplify the notation) for every $j \in \{1, \ldots, l\}$ we set

$$S_j := \mathrm{adv}_{\kappa\cdot j}\left((\rho, R(i+1), \rho')\right).$$

From the definition of a_{i+1} (cf. (21)) it follows easily that

$$|S_j| \le a_{i+1} \qquad (25)$$

for every j. Therefore (by (24))

$$\left|\mathrm{adv}_\kappa\left((\rho, R(i+1), \rho')\right)\right| \le a_{i+1}$$

and, since the choice of κ and ρ was arbitrary, (23) (and hence (15)) is proved. For the proof of (22) we will make use of the theory of *martingales* (see Definition 2 in Appendix A).

LEMMA 5. *$S_1, \ldots, S_l$ is a martingale difference sequence.*

PROOF. First, observe that for every j the bit $R(i+1, j+n-1)$ is independent of $(S_1, \ldots, S_{j-1})$. Let φ be a mapping that flips bit $j+n-1$ in an l-bit sequence $(q_1, \ldots, q_l) \in \{0,1\}^l$, i.e.,

$$\varphi((q_1, \ldots, q_l)) = (q_1, \ldots, q_{j+n-2}, \overline{q_{j+n-1}}, q_{j+n}, \ldots, q_l).$$

It is easy to see that for every k with prefix $\kappa \cdot j$ we have

$$c(k, (\rho, R(i+1), \rho') = -c(k, (\rho, \varphi(R(i+1)), \rho')$$

(in other words for the key k the function g guesses correctly for $(\rho, R(i+1), \rho')$ if and only if it guesses incorrectly for $(\rho, \varphi(R(i+1)), \rho'))$ and hence

$$\mathrm{adv}_{\kappa\cdot j}\left((\rho, R(i+1), \rho')\right) = -\mathrm{adv}_{\kappa\cdot j}\left((\rho, \varphi(R(i+1)), \rho')\right).$$

Therefore

$$P_{S_j|S_1\cdots S_{j-1}} = P_{-S_j|S_1\cdots S_{j-1}} \qquad (26)$$

which implies that for all values of $(s_1, \ldots, s_{j-1})$ we have $E[S_j \mid S_1 = s_1, \ldots, S_{j-1} = s_{j-1}] = 0$. □

Let us return to the proof of Lemma 3. From Lemma 8 (see Appendix A) and from the fact (25) that $|S_j| \le a_{i+1}$ we get

$$P\left(\left|\sum_{j=1}^{l} S_j\right| \ge \tau l a_{i+1}\right) \le 2e^{-l\tau^2/2}. \qquad (27)$$

Therefore by (24) we have

$$P\left(\left|\mathrm{adv}_\kappa\left((\rho, R(i+1), \rho')\right)\right| \ge \tau a_{i+1}\right) \le 2e^{-l\tau^2/2}. \qquad (28)$$

Note that (28) holds for any $\kappa \in \mathcal{K}_i$ and $\rho \in \mathcal{R}_i$, as our choice of these values was arbitrary. To finalize the proof we will show that (28) implies (22). For a better readability define, for every $\kappa \in \mathcal{K}_i$ and $\rho \in \mathcal{R}_i$, $D(\kappa, \rho)$ as the event that $|\mathrm{adv}_\kappa((\rho, R(i+1), \rho'))| \ge \tau a_{i+1}$. Let

$$\mathcal{D} := \{D(\kappa, \rho) : (\kappa, \rho) \in \mathcal{K}_i \times \mathcal{R}_i\} \qquad (29)$$

be the set of all such events. Also, set $\gamma := 2e^{-l\tau^2/2}$. Thus (28) can be rewritten as

$$\forall_{D \in \mathcal{D}}\ P(D) \le \gamma. \qquad (30)$$

Clearly (since $i < m$), to prove (22) it suffices to show that

$$P\left(\bigcup_{D \in \mathcal{D}} D\right) \le l^i 2^n \gamma. \qquad (31)$$

In order to prove (31) we will apply the union bound. More precisely, we will use the following simple fact (that holds for any $\mathcal{D}$ and γ such that (30) is satisfied):

$$P\left(\bigcup_{D \in \mathcal{D}} D\right) \le |\mathcal{D}|\,\gamma. \qquad (32)$$

What remains is to give a strong enough bound on the cardinality of $\mathcal{D}$. It is easy to see that a naive approach (i.e. just looking at the definition (29) of $\mathcal{D}$) will not work since $|\mathcal{K}_i \times \mathcal{R}_i| = n^i 2^{(l+n-1)i}$ is too large. Therefore we will use a somewhat more subtle argument. Let f_i denote the key-derivation function for the randomizers of height $m = i$. For given $\rho_0, \rho_1 \in \mathcal{R}_i$ such that

$$f^i(\kappa, \rho_0) = f^i(\kappa, \rho_1),$$

due to the definitions (2) of f and (13) of advantage, for every $\rho'' \in \mathcal{R}_{m-i}$ we have

$$\mathrm{adv}_\kappa\left((\rho_0, \rho'')\right) = \mathrm{adv}_\kappa\left((\rho_1, \rho'')\right),$$

and thus it is not difficult to see that the events $D(\kappa, \rho_0)$ and $D(\kappa, \rho_1)$ are equal. In other words, for every $\alpha \in \{0,1\}^n$ the set $\mathcal{D}_{\kappa,\alpha} := \{D(\kappa, \rho) : f^i(\kappa, \rho) = \alpha\}$ consists of at most one event. Since $|\mathcal{K}_i| = l^i$ and $\mathcal{D}$ is the union of all the sets $\mathcal{D}_{\kappa,\alpha}$, the set $\mathcal{D}$ contains at most $l^i 2^n$ distinct events. In other words, $|\mathcal{D}| \le l^i 2^n$. Combining this with (32) we get (31). This completes the proof □

PROOF OF LEMMA 2. Since $\mathrm{adv}(R) = A_0$, Lemmas 3 and 4 together imply

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

which implies (14) by changing the base in the right term from e to 2. □

5. ASYMPTOTICALLY OPTIMAL RANDOMNESS EFFICIENCY

Recall that Corollary 1 states that (roughly speaking) Alice and Bob can securely derive a key with randomness efficiency $\nu = 8\%$. A natural question to ask is for which maximal value of ν secure key-expansion is possible. In this section we prove that ν can be arbitrarily close to 1. We will again use the scheme defined in Section 3.1. For a given randomizer height m and width l, a length n of a derived key and a constant $\nu \in [0,1]$, let $M(m, l, n, \nu)$ be the expected bias of Eve's distribution of the derived key X (given her entire information), assuming that she applies the best (from her point of view) strategy. More precisely, let $M(m, l, n, \nu)$ be equal to

$$\max_{h:\{0,1\}^t \to \{0,1\}^{\lfloor \nu t \rfloor}} E\left[\beta_h(U, K)\right],$$

where $t = m(l + n - 1)$ and (cf. 3)

$$\beta_h(u, k) := d\left(P_{X|h(R)=u, K=k}\right).$$

To be more formal we need to be careful about how fast the parameters l, m and n grow with respect to each other. Let l and n be functions of m. First, we will assume that l is at least polynomial and at most exponential in m. Namely $l := \lambda(m)$, where λ is some fixed function such that for every m

$$\lambda(m) \geq m^3 \tag{33}$$

and

$$\lambda(m) \leq 2^m. \tag{34}$$

(The choice of the bounds (33) and (34) is rather arbitrary.) Second, we have to assume that the derived key is longer than the initial key. Therefore we fix an arbitrary constant $c > 1$ and assume that $n := cm \log_2 l$, which by (34) is at most cm^2. The main theorem of this section is as follows.

Theorem 2. *For every $\nu \in [0, 1), c > 1$ and λ as above, the value*

$$M(m, \lambda(m), cm \log_2(\lambda(m)), \nu)$$

decreases exponentially in m.

This result is primarily of theoretical interest since for ν close to 1 the parameter m must be quite large.

Let us first look at what we have proven so far (Theorem 1). It is not difficult to see that for all ξ and τ the value δ (defined by (5)) is at least $e^{-lm/2}$. Therefore for $s \geq lm(\log_2 e)/2$ the right hand side of (4) is at least n. Thus Theorem 1 gives no non-trivial bound on $E\left[\beta(U, K)\right]$ if ν exceeds $\frac{lm}{t}(\log_2 e)/2 \approx (\log_2 e)/2 \approx 0.72$.

Therefore we need to modify the main technical part. The factor $(\log_2 e)/2$ comes from Lemma 8 used in the proof of Lemma 3 (more precisely, it is used to obtain (27)). We will now prove an alternative inequality (Lemma 6). First, recall that (in the proof of Lemma 5) we in fact proved something stronger than the statement of the lemma. Namely we proved property (26).

Lemma 6. *Let $S_1, \ldots, S_l$ be a sequence of random variables such that for every j we have $|S_j| \leq a$ (for some $a \geq 0$) and $P_{S_j|S_1,\ldots,S_{j-1}} = P_{-S_j|S_1,\ldots,S_{j-1}}$. Then, for every $\tau \in [0.5, 1)$,*

$$P\left(\left|\sum_{j=1}^{m} S_j\right| \geq \tau la\right) \leq 2^{l(H(\tau)-1)+1}, \tag{35}$$

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

This lemma is incomparable with Lemma 8. On one hand it is weaker since it applies only when (26) is satisfied and for $\tau \geq 1/2$ (also, for τ not much greater than 1/2 it gives a worse bound). However, for our purposes (τ close to 1) it is stronger than Lemma 8.

Proof of Lemma 6. For every j set $T_j := |S_j|$ and

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

Clearly $S_j = T_j \cdot U_j$. Therefore, for every τ,

$$\sum_{j=1}^{l} S_j \geq \tau la$$

implies that $|\{j : U_j = 1\}| \geq \tau l$. It is not difficult to see that $U_1, \ldots, U_l$ are distributed independently and uniformly. Therefore the size of the set $\{j : U_j = 1\}$ is distributed according to the binomial distribution. Putting things together we get

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

The last step follows from the inequality

$$\sum_{j=0}^{\omega l} \binom{l}{j} \leq 2^{lH(\omega)} \tag{36}$$

which holds for all $\omega \in (0, 0.5]$ (in our case $\omega = 1 - \tau$). (For the proof of (36) see e.g. [18] Theorem 1.4.5, page 21.) Clearly by symmetry we also get

$$P\left(\sum_{j=1}^{l} S_j \leq -\tau la\right) \leq 2^{l(H(1-\tau)-1)}.$$

Thus (35) follows. □

If we now use Lemma 6 instead of Lemma 8 (in the proof of Lemma 3), then instead of (27) we get (35). This yields the following lemma that may be viewed as an alternative (and incomparable) version of Theorem 1.

Lemma 7. *For l, m, n, s, R, K and U as in Theorem 1, for any $\tau, \xi \in [0, 1]$ we have $E\left[\beta(U, K)\right] \leq n \cdot (2^s \delta' + \epsilon)$, where ϵ is defined by (6), and δ' by*

$$\delta' := 2^{m+(1-\xi)m(m\log_2 l+n+1)+lm(H(\tau)-1)(1-\xi)}.$$

We can now use Lemma 7 to prove Theorem 2.

Proof of Theorem 2. Fix some ν. From Lemma 7 we get that for every τ and ξ,

$$\begin{aligned} &M(m, \lambda(m), cm \log_2(\lambda(m)), \nu) \\ &\qquad \leq cm^2 \left(2^{\nu m(\lambda(m)+cm^2-1)} \delta' + \epsilon\right) \\ &\qquad \leq cm^2 \left(2^{cm^3+\nu m\lambda(m)} \delta' + \epsilon\right), \end{aligned} \tag{37}$$

where

$$\begin{aligned} \delta' &\leq 2^{m+(1-\xi)m\left(m^2+cm^2+1\right)+m\lambda(m)(H(\tau)-1)(1-\xi)} \\ &\leq 2^{c'm^3+m\lambda(m)(H(\tau)-1)(1-\xi)}. \end{aligned} \tag{38}$$

(for some constant c') and $\epsilon = \tau^{\xi m}/2$. It is easy to check that

$$\lim_{\tau \to 1} H(\tau) = 0.$$

Therefore there must exist $\tau, \xi \in (0,1)$ such that $(H(\tau) - 1)(1-\xi) \leq -\frac{1}{2}(\nu + 1)$. Fix such τ and ξ. Hence (from (38))

$$\delta' \leq 2^{c' m^3 - m\lambda(m)(\nu+1)/2}.$$

Thus (37) is at most

$$cm^2 \left(2^{-f(m)} + \epsilon\right), \qquad (39)$$

where $f(m) = \lambda(m)m(1-\nu)/2 - (c+c')m^3$. From the assumptions that $\nu < 1$ and (33) we get that $f(m) = \Omega(m^4)$. Therefore for sufficiently large m the value of $2^{-f(m)}$ is smaller than $\tau^{\xi m}/2$ and thus (39) is at most

$$cm^2 \cdot (2 \cdot \tau^{\xi m}/2) \leq cm^2 \tau^{\xi m},$$

which decreases exponentially in m. □

6. ACKNOWLEDGMENTS

We would like to thank Rasmus Pagh for a helpful discussion.

7. REFERENCES

[1] Y. Aumann, Y. Z. Ding, and M. O. Rabin. Everlasting security in the bounded storage model. To appear in IEEE Transactions on Information Theory, Apr. 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*, pages 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*, 5:2–28, 1992.

[5] 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)*, pages 493–502. IEEE Computer Society, 1998.

[6] 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*, pages 292–306. Springer-Verlag, 1997.

[7] T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley Series in Telecommunications. Wiley International, 1992.

[8] Y. Z. Ding. Oblivious transfer in the bounded storage model. In *Advances in Cryptology - CRYPTO 2001*, volume 2139, pages 155–170. Lecture Notes in Computer Science, 2001.

[9] Y. Z. Ding. *Provably Everlasting Security in the Bounded Storage Model.* PhD thesis, Harvard University, 2001. available at `http://www.deas.harvard.edu/~zong`.

[10] Y. Z. Ding and M. O. Rabin. Hyper-encryption and everlasting security. In *19th Annual Symposium on Theoretical Aspects of Computer Science*, volume 2285 of *Lecture Notes in Computer Science.* Springer-Verlag, 2002. To appear.

[11] D. Dubhashi and A. Panconesi. Concentration of measure for the analysis of randomised algorithms. draft available at `http://www.dsi.uniroma1.it/~ale`, Oct. 1998.

[12] S. Goldwasser and S. Micali. Probabilistic Encryption. *Journal of Computer and System Sciences*, 28:270–299, 1984.

[13] 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.

[14] U. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *Journal of Cryptology*, 5(1):53–66, 1992.

[15] U. Maurer. Secret key agreement by public discussion. *IEEE Transaction on Information Theory*, 39:733–742, 1993.

[16] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[17] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

[18] J. H. van Lint. *Introduction to Coding Theory.* Springer-Verlag, 3rd edition, 1998.

APPENDIX

A. MARTINGALES

In this section we sketch the basic ideas behind the theory of martingales that we use in the proof of Lemma 5. We refer to [11] and [16] for more on this subject.

Definition 2. *A* martingale difference sequence *is a sequence of real-valued random variables $S_1, \ldots, S_l$ such that for every j and every $s_1, \ldots, s_{j-1}$ we have*

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

The following lemma follows directly from Azuma's inequality (see e.g. Theorem 4.16 in [16], page 92, or [3] for the original result).

Lemma 8. *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 $\tau \in \mathbf{R}$,*

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

(To derive Lemma 8 from Theorem 4.16 in [16] 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 $c_i := a$.)

B. PROOF OF LEMMA 1

In the proof we will use the following simple fact that holds for every sequence of random variables $X_1, \ldots, X_j$.

$$\mu_j(P_X) := \tfrac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \left|P_{X_1,\ldots,X_j}(\mathbf{x}, 0) - P_{X_1,\ldots,X_j}(\mathbf{x}, 1)\right|.$$

We show inductively (over j) that

$$d(P_{X_1 \cdots X_j}) \leq \sum_{i=1}^{j} \mu_i(P_X),$$

for every $j \in \{1, \ldots, n\}$. Case $j = 1$ is trivial. Assume that the hypothesis holds for some $j - 1$. To avoid too many subscripts let Q be the probability distribution $P_{X_1 \cdots X_j}$. Clearly we have that $d(P_{X_1 \cdots X_{j-1}}) = \frac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \left| Q(\mathbf{x}, 0) + Q(\mathbf{x}, 1) - 2^{-j+1} \right|$. Therefore

$$\begin{aligned}
\sum_{i=1}^{j} \mu_i(P_X) &\geq d(P_{X_1,\ldots,X_{j-1}}) + \mu_j(P_X) && (40) \\
&= \tfrac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \left| Q(\mathbf{x}, 0) + Q(\mathbf{x}, 1) - 2^{-j+1} \right| \\
&\qquad + \tfrac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \left| Q(\mathbf{x}, 0) - Q(\mathbf{x}, 1) \right| \\
&= \tfrac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \Big(\left| Q(\mathbf{x}, 0) + Q(\mathbf{x}, 1) - 2^{-j+1} \right| \\
&\qquad + \left| Q(\mathbf{x}, 0) - Q(\mathbf{x}, 1) \right| \Big) \\
&= \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \max \left(\left| Q(\mathbf{x}, 0) - 2^{j} \right|, \left| Q(\mathbf{x}, 1) - 2^{j} \right| \right) && (41) \\
&\geq \tfrac{1}{2} \sum_{\mathbf{x} \in \{0,1\}^{j-1}} \left(\left| Q(\mathbf{x}, 0) - 2^{j} \right| + \left| Q(\mathbf{x}, 1) - 2^{j} \right| \right) && (42) \\
&= d(Q) = d(P_{X_1,\ldots,X_j})
\end{aligned}$$

where (40) follows from the induction hypothesis, (41) follows from the fact that $\max(|a|, |b|) = \frac{1}{2}(|a+b| + |a-b|)$, for every $a, b \in \mathbf{R}$, and (42) follows from $\max(a, b) \geq \frac{1}{2}(a + b)$. □