In *Proceedings of CRYPTO '97* (B. Kaliski, ed.), Lecture Notes in Computer Science, Springer, 1997.

Unconditional Security Against Memory-Bounded Adversaries

Christian Cachin* Ueli Maurer

Department of Computer Science
ETH Zürich
CH-8092 Zürich, Switzerland
cachin@acm.org maurer@inf.ethz.ch

May 26, 1997

Abstract

We propose a private-key cryptosystem and a protocol for key agreement by public discussion that are unconditionally secure based on the sole assumption that an adversary's memory capacity is limited. No assumption about her computing power is made. The scenario assumes that a random bit string of length slightly larger than the adversary's memory capacity can be received by all parties. The random bit string can for instance be broadcast by a satellite or over an optical network, or transmitted over an insecure channel between the communicating parties. The proposed schemes require very high bandwidth but can nevertheless be practical.

1 Introduction

One of the most important properties of a cryptographic system is a proof of its security under reasonable and general assumptions. However, every design involves a trade-off between the strength of the security and further important qualities of a cryptosystem, such as efficiency and practicality.

The security of all currently used cryptosystems is based on the difficulty of an underlying computational problem, such as factoring large numbers or computing discrete logarithms in the case of many public-key systems. Security proofs for these systems show that the ability of the adversary to defeat the cryptosystem with significant probability contradicts the assumed difficulty of the problem [24]. Although the hardness of these problems is unquestioned at the moment, it can be dangerous to base the security of the global information economy on a very small number of mathematical problems. Recent advances in quantum computing show that precisely these two problems, factoring and discrete logarithm, could be solved efficiently if quantum computers could be built [27].

An alternative to proofs in the computational security model is offered by the stronger notion of information-theoretic or *unconditional* security where no limits on an adversary's computational power are assumed. The first information-theoretic definition of perfect secrecy by Shannon [26] led immediately to his famous impracticality theorem, which states, roughly, that the shared secret key in any perfectly secure cryptosystem must be at least as long as the plaintext to be encrypted. Vernam's one-time pad is the prime example of a perfectly secure

*Current address: MIT Laboratory for Computer Science, 545 Technology Square, Cambridge MA 02139.

but impractical system. Unconditional security was therefore considered too expensive for a long time.

However, recent developments show how Shannon's model can be modified [16] to make practical provably secure cryptosystems possible. The first modification is to relax the requirement that perfect security means complete independence between the plaintext and the adversary's knowledge and to allow an arbitrarily small correlation. The second, crucial modification removes the assumption that the adversary receives exactly the same information as the legitimate users. The following two primitives are perhaps the most realistic mechanisms proposed so far for limiting the information available to the adversary.

Quantum Channel: Quantum cryptography was developed mainly by Bennett and Brassard during the 1980's [3]. It makes use of photons, i.e. polarized light pulses of very low intensity, that are transmitted over a fiber-optical channel. In the basic quantum key agreement protocol, this allows two parties to generate a secret key by communicating about the received values. The unconditional secrecy of the key is guaranteed by the uncertainty principle of quantum mechanics. Current implementations of quantum key distribution span distances of 20–30 kilometers.

Noisy Channel: In this model proposed by Maurer, the output of a random source is transmitted to the participants over partially independent noisy channels that insert errors with certain probabilities [19]. Two parties can then generate a secret key from their received values by public discussion. The secrecy of the key is based on the information differences between the channel outputs and on the assumption that no channel is completely error-free. This system is practical because it works also in the realistic case where the adversary receives the random source via a much better channel than the legitimate users. The power of a noisy channel was also demonstrated by Crépeau and Kilian who showed that unconditionally secure bit commitment and oblivious transfer can be based on this primitive [11, 10].

In this paper, we show how to realize unconditionally secure encryption based on a third assumption: a limit on the memory size of the adversary. This means that an enemy can use unlimited computing power to compute any probabilistic function of some huge amount of public data, which is infeasible to store. As long as the function's output size does not exceed the number of available storage bits, we can prove that the proposed private-key system and public key agreement protocol are information-theoretically secure from this sole assumption.

The public data is the output of a random source that is broadcast at very high rate. The legitimate users Alice and Bob randomly select a small subset of the broadcast each and store these values. (How this selection is performed will be described below.) Because of the random selection process, the average fraction of the information of an adversary Eve about the selected subset is roughly the same as her fraction of information about the complete broadcast. By applying privacy amplification [2], Alice and Bob can then eliminate Eve's partial knowledge about the selected subset. (The random source does not have to be independent from the users, e.g. Alice could produce the random data herself and transmit it to Bob over a public channel.)

We describe how two different cryptographic tasks can be implemented using this mechanism, depending on how Alice and Bob select the random subset. First, if they share a short secret key initially that can be used to select identical subsets, the system realizes *private-key encryption*. Second, even if Alice and Bob do not share any secret information at the beginning, they can perform a *key agreement protocol* by public discussion: They select and store independently a random subset of the broadcast data. After some predetermined interval they publicly exchange the indices of their selected positions and determine the positions contained in both subsets. Privacy amplification is applied to the part of the broadcast they have in common.

Our model seems realistic because current communication and high-speed networking technologies allow broadcasting at rates of multiple gigabits per second. Storage systems that are hundreds of terabytes large, on the other hand, require a major investment by a potential adversary. Although this is within reach of government budgets, for example, the method is attractive for the following three reasons: First, the security can be based only on the assumption about the adversary's memory capacity. Second, storage costs scale linearly and can therefore be estimated accurately. Third, the system offers 'proactive' security in the sense that a future increase in storage capacity cannot break the secrecy of messages encrypted earlier.

A precursor of this system is the Rip van Winkle cipher proposed by Massey and Ingemarsson [17, 16]. This private-key system is provably computationally secure but totally impractical because a legitimate receiver must wait even longer for receiving a message than it takes an adversary to decrypt it.

Related to our work is a paper by Maurer [18] that describes a system based on a large public randomizer which cannot be read entirely within feasible time. Maurer's paper contains also the idea of realizing provably secure encryption based only on assumptions about an enemy's available memory. Such a system for key agreement was described by Mitchell [20], but without security proof. Our analysis provides the first proof that unconditional security can be achieved against memory-bounded adversaries. (Recently, Aumann and Rabin [22] proved a conjecture of Maurer's paper [18] with the same effect.)

We borrow some methods from the work of Zuckerman and others on so-called extractors of uniform randomness from weak random sources [29]. Extractors are tools developed for running randomized algorithms with non-perfect randomness instead of uniform random bits. Nisan [21] presents a highly readable introduction to extractors and a survey of their applications.

The paper is organized as follows. After reviewing some information-theoretic concepts in Section 2, we introduce the building blocks of our system in Section 3. Our main result concerning Eve's information about the randomly selected subset is given in Section 4. Sections 5 and 6 describe how to realize private-key encryption and public key agreement, respectively. The paper concludes with a discussion of the underlying assumptions and future perspectives.

2 Preliminaries

We assume that the reader is familiar with the notion of entropy and the basic concepts of Shannon's information theory [9]. We repeat some fundamental definitions in this section and introduce the notation. All logarithms in this paper are to the base 2. The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

A random variable X induces a probability distribution P_X over an alphabet $\mathcal{X}$. Random variables are denoted by capital letters. If not stated otherwise, the alphabet of a random variable is denoted by the corresponding script letter. A sequence $X_1, \ldots, X_n$ of random variables with the same alphabet is denoted by X^n.

The *(Shannon) entropy* of a random variable X with probability distribution P_X and alphabet $\mathcal{X}$ is defined as

$$H(X) \;=\; -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).$$

The *binary entropy function* is $h(p) = -p \log p - (1-p) \log(1-p)$. The *conditional entropy* of X conditioned on a random variable Y is

$$H(X|Y) \;=\; \sum_{y \in \mathcal{Y}} P_Y(y) H(X|Y=y)$$

where $H(X|Y=y)$ denotes the entropy of the conditional probability distribution $P_{X|Y=y}$. The *mutual information* of X and Y is the reduction of the uncertainty of X when Y is learned:

$$I(X;Y) = H(X) - H(X|Y).$$

The *variational distance* between two probability distributions P_X and P_Y over the same alphabet $\mathcal{X}$ is

$$\|P_X - P_Y\|_v \;=\; \max_{\mathcal{X}_0 \subseteq \mathcal{X}} \Big| \sum_{x \in \mathcal{X}_0} P_X(x) - P_Y(x) \Big| \;=\; \frac{1}{2} \sum_{x \in \mathcal{X}} \Big| P_X(x) - P_Y(x) \Big|.$$

$\|P_X - P_Y\|_v \leq \epsilon$ implies that X behaves like Y except with probability at most ϵ, i.e., any property of X is shared by Y with probability at least $1-\epsilon$.

The *Rényi entropy of order* α of a random variable X with alphabet $\mathcal{X}$ is

$$H_\alpha(X) \;=\; \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha$$

for $\alpha \geq 0$ and $\alpha \neq 1$ [23]. Because the limiting case of Rényi entropy for $\alpha \to 1$ is Shannon entropy, we can extend the definition to $H_1(X) = H(X)$. In the other limiting case $\alpha \to \infty$, we obtain the *min-entropy*, defined as

$$H_\infty(X) = -\log \max_{x \in \mathcal{X}} P_X(x).$$

For a fixed random variable X, Rényi entropy is a continuous positive decreasing function of α. For $0 < \alpha < \beta$,

$$H_\alpha(X) \;\geq\; H_\beta(X)$$

with equality if and only if X is the uniform distribution over $\mathcal{X}$ or over a subset of $\mathcal{X}$. In particular, $\log|\mathcal{X}| \geq H_\alpha(X) \geq 0$ for $\alpha \geq 0$ and $H(X) \geq H_\alpha(X)$ for $\alpha > 1$.

3 Pairwise Independence and Entropy Smoothing

This section contains a short review of entropy smoothing with universal hashing. We start by repeating the construction of a sequence of pairwise independent random variables using universal hash functions.

Universal hash functions were introduced by Carter and Wegman [8, 28] and have found many applications in theoretical computer science [15]. A 2-*universal hash function* is a set $\mathcal{G}$ of functions $\mathcal{X} \to \mathcal{Y}$ if, for all distinct $x_1, x_2 \in \mathcal{X}$, there are at most $|\mathcal{G}|/|\mathcal{Y}|$ functions g in $\mathcal{G}$ such that $g(x_1) = g(x_2)$.

A *strongly* 2-*universal hash function* is a set $\mathcal{G}$ of functions $\mathcal{X} \to \mathcal{Y}$ if, for all distinct $x_1, x_2 \in \mathcal{X}$ and all (not necessarily distinct) $y_1, y_2 \in \mathcal{Y}$, exactly $|\mathcal{G}|/|\mathcal{Y}|^2$ functions from $\mathcal{G}$ take x_1 to y_1 and x_2 to y_2.

A strongly 2-universal hash function can be used to generate a sequence of pairwise independent random variables in the following way: Select $G \in \mathcal{G}$ uniformly at random and apply it to any fixed sequence $x_1, \ldots, x_l$ of distinct values in $\mathcal{X}$. Let $Y_j = G(x_j)$ for $j = 1, \ldots, l$. It can easily be verified that $Y_1, \ldots, Y_l$ are pairwise independent and uniformly distributed random variables over $\mathcal{Y}$. The advantage of this technique, compared to selecting n independent samples of Y, is that it requires only $2\log|\mathcal{Y}|$ instead of $n \log|\mathcal{Y}|$ random bits.

An often-used example for a strongly 2-universal hash function from $GF(2^n)$ to $GF(2^m)$ is the set

$$\mathcal{G} \;=\; \big\{ g(x) = \mathrm{msb}_m(a_1 x + a_0) \;\big|\; a_0, a_1 \in GF(2^n) \big\}$$

where $\mathrm{msb}_m(x)$ denotes the m most significant bits of x and $a_1x + a_0$ is computed in $GF(2^n)$. This construction has the nice property that when $\mathcal{G}$ is used to generate a sequence of pairwise independent random variables, all values in the sequence are distinct if and only if $a_1 \neq 0$. We will assume that $a_1 \neq 0$ whenever the pairwise independence construction is used and refer to the resulting distribution as "uniform and pairwise independent" although repeating values are excluded.

The strongly 2-universal family $\mathcal{G}$ is 2-universal even if a_0 is always set to 0. Thus, a member of the 2-universal family can be specified with only n bits.

2-universal hash functions are also the main technique to concentrate the randomness inherent in a probability distribution by a result known in different contexts as Entropy Smoothing Theorem, Leftover Hash Lemma [14], or Privacy Amplification Theorem [2].

In cryptography, privacy amplification is used to extract a short secret key from shared information about which an adversary has partial knowledge. Assume Alice and Bob share a random variable W, while an eavesdropper Eve knows a correlated random variable V that summarizes her knowledge about W. The details of the distribution P_{WV}, and thus of Eve's information V about W, are unknown to Alice and Bob, except that they assume a lower bound on the Rényi entropy of order 2 of $P_{W|V=v}$ for the particular value v that Eve observes.

Using a public channel, which is susceptible to eavesdropping but immune to tampering, Alice and Bob wish to agree on a function g such that Eve knows nearly nothing about $g(W)$. The following theorem by Bennett, Brassard, Crépeau, and Maurer [2] shows that if Alice and Bob choose g at random from a universal hash function $\mathcal{G} : \mathcal{W} \to \mathcal{Y}$ for suitable $\mathcal{Y}$, then Eve's information about $Y = g(W)$ is negligible.

Theorem 1 ([2]). *Let X be a random variable over the alphabet $\mathcal{X}$ with Rényi entropy $H_2(X)$, let G be the random variable corresponding to the random choice (with uniform distribution) of a member of a 2-universal hash function $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$, and let $Y = G(X)$. Then*

$$H(Y|G) \;\geq\; \log|\mathcal{Y}| - \frac{2^{\log|\mathcal{Y}| - H_2(X)}}{\ln 2}. \tag{1}$$

To apply the theorem in the described scenario, replace P_X by the conditional probability distribution $P_{W|V=v}$. Cachin has recently extended the theorem to Rényi entropy of order α for any $\alpha > 1$ [6].

4 Extracting a Secret Key from a Randomly Selected Subset

We are now going to show how and why Alice and Bob can exploit the fact that an adversary Eve cannot store the complete output of a public random source that is broadcast to the participants. The security proof consists of three steps. In the first step, we use the fact that Eve's storage capacity is limited to establish a lower bound on the min-entropy of Eve about the broadcast bits. The second step shows that Eve's min-entropy about a randomly selected subset of the broadcast bits is large with high probability. In the third step, we apply privacy amplification to the selected subset to obtain the secret key.

Suppose the output of a uniformly distributed binary source R is broadcast over an error-free channel and can be received by all participants. The source can be independent from the participants or it can be operated by one of the legitimate users, e.g. Alice can generate R and transmit it over an authenticated public channel to Bob. More generally, any source that is trusted to output random bits and has a sufficient capacity can be used. The channel must have high capacity, which could be realized, for example, using satellite technology for digital TV broadcasting or all-optical networks. The channel is used n times in succession and the broadcast bits are denoted by $R^n = R_1, \ldots, R_n$. We assume that Eve has a total of $m < n$

storage bits available and therefore cannot record the complete broadcast, leaving her only with partial knowledge about R^n.

During the broadcast, Eve may compute an arbitrary function of R^n with unlimited computing power and can also use additional private random bits. We model the output of the function to be stored in her m bits of memory by the random variable Z with alphabet $\mathcal{Z}$, subject to $\log|\mathcal{Z}| \leq m$.

Because R^n is uniformly distributed, its Rényi entropy of any order $\alpha \geq 0$ and its Shannon entropy satisfy $H_\alpha(R^n) = H(R^n) = H_\infty(R^n) = n$. The following lemma shows that the min-entropy of R^n given Z, which corresponds to Eve's knowledge about R^n, is at least $n - m$ for all but a negligible fraction of the values of Z. More precisely, the lemma implies that for any $r > 0$, the particular value z that Z takes on satisfies $H_\infty(R^n|Z = z) \geq n - m - r$, except with probability at most 2^{-r}.

Lemma 2. *Let X be a random variable with alphabet $\mathcal{X}$, let Z be an arbitrary random variable with alphabet $\mathcal{Z}$, and let $r > 0$. Then with probability at least $1 - 2^{-r}$, Z takes on a value z for which*

$$H_\infty(X|Z = z) \;\geq\; H_\infty(X) - \log|\mathcal{Z}| - r.$$

Proof. Let $p_0 = 2^{-r}/|\mathcal{Z}|$. Thus, $\sum_{z:P_Z(z)<p_0} P_Z(z) < 2^{-r}$. It follows for all z with $P_Z(z) \geq p_0$

$$\begin{aligned}
H_\infty(X|Z = z) &= -\log \max_{x\in\mathcal{X}} P_{X|Z=z}(x) \\
&= -\log \max_{x\in\mathcal{X}} \frac{P_X(x)P_{Z|X=x}(z)}{P_Z(z)} \\
&\geq -\log \max_{x\in\mathcal{X}} \frac{P_X(x)}{p_0} \\
&= H_\infty(X) - r - \log|\mathcal{Z}|
\end{aligned}$$

which proves the lemma. □

For the rest of this section, we denote Eve's knowledge of R^n, given the particular value $Z = z$ she observed, by the random variable $X^n = X_1, \ldots, X_n$ with alphabet $\mathcal{X}^n = \{0,1\}^n$. The distribution of X^n is arbitrary and only subject to $H_\infty(X^n) \geq n - m - r$ by Lemma 2.

The strategy of the legitimate users Alice and Bob is to select the values at l positions

$$\mathbf{S} \;=\; [S_1, \ldots, S_l] \quad \text{with} \quad S_1, \ldots, S_l \in \{1, \ldots, n\}$$

randomly from the broadcast symbols X^n. $\mathbf{S}$ is a vector-valued random variable taking on values $\mathbf{s} \in \{1, \ldots, n\}^l$ and the list of selected positions $X_{S_1}, \ldots, X_{S_l}$ is denoted by $X^{\mathbf{S}}$. Because this selection is performed with uniform distribution according to the pairwise independence construction of a sequence of l values from $\{1, \ldots, n\}$ as described in Section 3, the resulting $S_1, \ldots, S_l$ are all distinct and $\mathbf{S}$ can also be viewed as a set of l values. In addition, $\mathbf{S}$ can be determined efficiently from $2\log n$ bits.

We assume that the value of $\mathbf{S}$ is known whenever the random variable $X^{\mathbf{S}}$ is used. In the private-key system described later, Eve is thus supposed to obtain $\mathbf{S}$ from an oracle *after* the public random string is broadcast.

How much does Eve know about the bits selected by Alice and Bob? Intuitively, one would expect that the fraction of bits in $X^{\mathbf{S}}$ known to Eve corresponds to the fraction of bits in X^n that Eve knows (here a bit is not to be understood as a binary digit, but in the information-theoretic sense). This is indeed the case, as was observed before by Zuckerman and others in the context of weak random sources [29, 21]. It is easy to show that the fraction of Eve's Shannon information corresponds to the expected value [5, Theorem 5.10]. However, privacy amplification can only be applied if a lower bound on the Rényi entropy of order 2 of $X^{\mathbf{S}}$ is known, which follows from the stronger bound on the min-entropy by Lemma 3.

The cited proof for Shannon information works only because Shannon entropy has the intuitive property that side information can only reduce the average uncertainty. This is not the case for expected conditional Rényi entropy of order $\alpha > 1$ and is the main obstacle for extending the proof to Rényi entropy. However, the following stronger result by Zuckerman [29] shows that also the fraction of Eve's min-entropy in the selected positions is, with high probability, close to the corresponding fraction of the total min-entropy. Because the min-entropy of a random variable is a lower bound for its Rényi entropy for any $\alpha > 0$, the lemma is sufficient for applying privacy amplification to the selected subset.

Lemma 3. *Let X^n be a random variable with alphabet $\{0,1\}^n$ and min-entropy $H_\infty(X^n) \geq \delta n$ (where $\frac{1}{n} \leq \delta \leq 0.9453$), let $\mathbf{S} = [S_1, \ldots, S_l]$ be chosen pairwise independently as described in Section 3, let $\rho \in [0, \frac{1}{3}]$ be such that $h(\rho) + \rho \log \frac{1}{\delta} + \frac{1}{n} = \delta$, and let $\epsilon = \sqrt{4/(\rho l) + 2^{\rho n \log \delta}}$. Then, for every value $\mathbf{s}$ of $\mathbf{S}$ there exists a random variable $A^l(\mathbf{s})$ with alphabet $\{0,1\}^l$ and min-entropy $H_\infty(A^l(\mathbf{s})) \geq \rho l/2$ such that with probability at least $1-\epsilon$ (over the choice of $\mathbf{S}$), $P_{X^{\mathbf{S}}}$ is ϵ-close to $P_{A^l(\mathbf{S})}$ in variational distance, i.e.*

$$\forall \mathbf{s} \, : \, \exists A^l(\mathbf{s}) \, : \, \mathrm{P}\big[\, \| P_{X^{\mathbf{S}}} - P_{A^l(\mathbf{S})} \|_v \leq \epsilon \,\big] \;\geq\; 1 - \epsilon.$$

Remark. For fixed, large n, the value of ρ resulting from the choice in the lemma increases monotonically with δ and for δ smaller than about 0.9453 there always exists a unique $\rho \in [0, \frac{1}{3}]$ satisfying $h(\rho) + \rho \log \frac{1}{\delta} = \delta$, as can be verified easily.

Proof. The statement of the lemma is slightly different from Zuckerman's asymptotic result [29, Lemma 9] with respect to ρ (that we use in place of α) and ϵ, but follows also from the original proof. We describe here only the differences that lead to our formulation of the lemma.

It is straightforward to verify that $\binom{n}{i-1} = \frac{i}{n-i+1}\binom{n}{i}$ and therefore $\binom{n}{i-1} < \frac{1}{2}\binom{n}{i}$ for $i \leq n/3$. This implies $\binom{n}{i-j} < 2^{-j}\binom{n}{i}$ for $i \leq n/3$ and $0 \leq j \leq i$, from which the bound

$$\sum_{i=0}^{k} \binom{n}{i} \;<\; \binom{n}{k} \sum_{i=0}^{k} 2^{-i} \;<\; 2\binom{n}{k} \tag{2}$$

for any $k \leq n/3$ follows immediately. The approximation of $\binom{n}{i}$ by the binary entropy function [9], $\frac{1}{n+1} 2^{nh(\frac{i}{n})} \leq \binom{n}{i} \leq 2^{nh(\frac{i}{n})}$, implies

$$2 \cdot 2^{nh(\rho)} \;\geq\; 2\binom{n}{\lfloor \rho n \rfloor} \;>\; \sum_{i=0}^{\lfloor \rho n \rfloor} \binom{n}{i},$$

where the second inequality follows from (2) for $\rho \leq \frac{1}{3}$. Thus choosing ρ as described in the statement of the lemma guarantees that

$$2^{-\delta n} \cdot \sum_{i=0}^{\lfloor \rho n \rfloor} \binom{n}{i} \;<\; 2^{-\delta n} \cdot 2^{nh(\rho)+1} \;=\; 2^{-\rho n \log \frac{1}{\delta}}$$

as required in the proof of Lemma 12 in [29]. The choice of ϵ is the value resulting at the end of the proof of Lemma 10 in [29]. □

We are now ready to state the main result of this section. First, we summarize the scenario and the choice of the parameters.

Let R^n be a random n-bit string with uniform distribution that is broadcast to Alice and Bob who want to generate a secret key and to the adversary Eve who has a total of $m < n$ bits of memory available. Let the random variable Z denote Eve's knowledge about R^n, let $\varepsilon_1, \varepsilon_2 > 0$ be arbitrary error probabilities, and let $\Delta > 0$ be a parameter that denotes the amount of information that may leak to Eve. Let the parameters

1. $\delta = \min\{0.9453, \frac{1}{n}(n - m - \log \frac{1}{\varepsilon_1})\}$;
2. ρ such that $h(\rho) + \rho \log \frac{1}{\delta} + \frac{1}{n} = \delta$;
3. $l = \lfloor (\rho\varepsilon_2^2 - \rho 2^{-\rho n \log \frac{1}{\delta} - 2})^{-1} \rfloor$;
4. $r = \lfloor \log \Delta + \rho l/2 - 1 \rfloor$.

Alice and Bob select $\mathbf{S} = [S_1, \ldots, S_l]$ randomly from $\{1, \ldots, n\}$ with the pairwise independence construction as described in Section 3 and store the bits $R^{\mathbf{S}} = R_{S_1}, \ldots, R_{S_l}$. Then they select a function $G \in \mathcal{G}$ uniformly at random from a 2-universal hash function $\mathcal{G}$ from l-bit strings to r-bit strings and compute $K = G(R^{\mathbf{S}})$ as their secret key. The random experiment consists of the choices of R^n, Z, $\mathbf{S}$, and G. As mentioned before, the theorem is proved under the (weaker) assumption that $\mathbf{S}$ is known to Eve, although this may not even be the case.

Theorem 4. *In the described scenario, there exists a security event $\mathcal{E}$ that has probability at least $1 - \varepsilon_1 - \varepsilon_2$ such that Eve's information about K, given G, given her particular knowledge $Z = z$ about R^n, given $\mathbf{S} = \mathbf{s}$, and given $\mathcal{E}$, is at most Δ. Formally,*

$$\exists \mathcal{E} \;:\; \mathrm{P}[\mathcal{E}] \geq 1 - \varepsilon_1 - \varepsilon_2 \quad \textit{and} \quad I(K; G | Z = z, \mathbf{S} = \mathbf{s}, \mathcal{E}) \;\leq\; \Delta.$$

Proof. Applying Lemma 2 with error probability ε_1 shows that

$$H_\infty(R^n | Z = z) \;\geq\; n - m - \log \frac{1}{\varepsilon_1},$$

leading to the value of δ. Lemma 3 shows that $\mathbf{S}$ takes on a value $\mathbf{s}$ such that there is a distribution $P_{A^l(\mathbf{s})}$ within $\varepsilon_2/2$ of $P_{R^{\mathbf{s}}|Z=z}$ with probability $1 - \varepsilon_2/2$. Privacy amplification can be applied because

$$H_2(A^l(\mathbf{s})) \;\geq\; H_\infty(A^l(\mathbf{s})) \;\geq\; \rho l/2.$$

The choice of r guarantees $H(K|G, Z = z, \mathbf{S} = \mathbf{s}) \geq r - \Delta$ by Theorem 1 because

$$2^{r - H_2(A^l(\mathbf{s}))} / \ln 2 \;\leq\; \Delta.$$

Failure of the uniformity bound, which is equivalent to the event $\overline{\mathcal{E}}$, consists of the union of the following three events. First, the bound of Lemma 2 can fail with probability at most ε_1. Second, $A^l(\mathbf{s})$ may deviate from the random variable with distribution $P_{R^{\mathbf{s}}|Z=z}$ with probability at most $\varepsilon_2/2$ and third, an $\mathbf{s}$ such that the distance $\|P_{X^{\mathbf{s}}} - P_{A^l(\mathbf{s})}\|_v$ is outside of the allowed range occurs with probability at most $\varepsilon_2/2$ in Lemma 3. Applying the union bound, we see that $\mathrm{P}[\mathcal{E}] \geq 1 - \varepsilon_1 - \varepsilon_2$ and

$$H(K|G, Z = z, \mathbf{S} = \mathbf{s}, \mathcal{E}) \;\geq\; r - \Delta.$$

The theorem now follows from the definition of mutual information upon noting that $H(K|Z = z, \mathbf{S} = \mathbf{s}, \mathcal{E}) \leq r$. □

In a realistic cryptographic application of Theorem 4, the choice of the parameters is somewhat simplified because m is typically very large and because choosing a reasonable safety margin implies $n \gg m$. In this case, the parameters are $\delta = 0.9453$ and $\rho = \frac{1}{3}$, and l depends almost only on ε_2 and is close to $3/\varepsilon_2^2$. Thus, the storage required by Alice and Bob and the size of the resulting secret key are inverse proportional to the square of the desired error probability.

5 A Private-Key System

We now describe an example of a practical private-key encryption system that offers virtually the same security as the one-time pad. Assume Alice and Bob share a secret key K_0 and have both access to the broadcast public random source R^n. In addition, they are connected by an authenticated public channel on which Eve can read but not modify messages. For the pairwise independent selection of $\mathbf{S}$, the size of K_0 must be $2 \log n$ bit. However, no initial communication between the partners is needed because the interval to observe R and other parameters like l, r, ε_1, and ε_2 are fixed. The authenticated public channel is needed to exchange the description of the hash function G, which is used to extract the secret value K from $R^{\mathbf{S}}$.

In a straightforward implementation, Alice and Bob need $l(\log n + 1)$ bit of storage to hold $\mathbf{S} = [S_1, \ldots, S_l]$ and the values of $R^{\mathbf{S}}$. Because R^n is broadcast at high speed, the positions to observe must be precomputed and be recalled in increasing order. The legitimate users must only be able to synchronize on the broadcast channel and to read one bit from time to time. An adversary, however, needs equipment with high bandwidth from the channel interface through to mass storage in order to store a substantial part of R^n.

The following considerations demonstrate that this system is on the verge of being practical. The broadcast channel could be realized by a satellite. Typically, current communications satellites have a capacity of 1–10 Gbit/s [25]. Commercial satellite communications services offer broadcast data rates up to 0.8 Gbit/s at consumer electronics prices. Far more capacity is offered by fiber optical networks [12]. The test bed of the All-Optical Networking Consortium, for example, has a capacity of 1 Tbit/s and has been demonstrated at 130 Gbit/s (which was only limited by the number of sources available). On the other hand, tape libraries with capacities in the PByte range (1 PetaByte = 2^{50} or about 10^{15} bytes) are a major investment [13].

As an example for the private-key system, consider a 16 Gbit/s satellite channel that is used for one day, making $n = 1.5 \times 10^{15}$. The size of the secret key K_0 is only 102 bit. Assume the adversary can store 100 TByte ($m = 8.8 \times 10^{14}$). Using $\Delta = 10^{-20}$ and error probabilities $\varepsilon_1 = 10^{-20}$ and $\varepsilon_2 = 10^{-4}$, we see that $\delta = 0.41$, $\rho = 0.060$, $l = 1.7 \times 10^9$, and $r = 5.0 \times 10^7$, that is, about 6.0 MByte of virtually secret information K can be extracted. The legitimate users need only 10 GByte of storage each to hold the indices and the selected bits. For privacy amplification, one of them has to announce the randomly chosen universal hash function, which takes about 197 MByte. An adversary knows not more than 10^{-20} bit of K except with probability about 10^{-4}. K can be used directly for encryption with a one-time pad, for example.

The memory requirements of Alice and Bob can be reduced if fast computation enables an implicit representation of the indices $\mathbf{S}$. This seems feasible because only simple operations are needed for the pairwise independence selection method. Assuming for example that the l values can be computed in one minute, only the positions to be observed within the next minute must be stored. With the figures of the preceding example, this reduces the storage requirements to only 7 MByte for the current block of indices plus a total of 197 MByte for $R^{\mathbf{S}}$. If the computation of the indices takes longer, observation of the random broadcast could also be halted until the indices are available.

The system can be used repeatedly with only one initial key K_0, because a small part of the secret key K obtained in the first round can be used safely as the secret key for the subsequent round and so forth. In addition, some part of K can be employed to relax the authenticity requirement for the public channel using unconditionally secure message authentication techniques [28].

6 Key Agreement by Public Discussion

Our methods can also be used to establish a secret key between two users not sharing secret information who have access to the random broadcast and are linked by a public channel. Communication on the public channel is assumed to be authenticated, i.e. the adversary can read but not modify messages. This system offers public key agreement with virtually the same security as the one-time pad under the sole assumption that the adversary's memory capacity is limited. (The public communication channel is different from the public broadcast channel whose only purpose is to distribute a large number of random bits.)

To agree on a secret key, Alice and Bob independently select and store a subset of the broadcast random bits R^n. After a predetermined amount of time, they announce the chosen set of positions on the public channel. The secret key can then be extracted from the values of R^n at the common positions using privacy amplification. To keep the communication and storage requirements for Alice and Bob at a reasonable level, it is crucial that they use a memory-efficient description of the index set. Fortunately, the pairwise independent selection method achieves this.

Both Alice and Bob select a sequence of q uniform and pairwise independent indices $T_1, \ldots, T_q$ and $U_1, \ldots, U_q$, respectively, from $\{1, \ldots, n\}$ as described in Section 3. (The values of q and the other parameters $n, l, r, \varepsilon_1, \varepsilon_2$ are fixed and also known to Eve.) Alice stores the values of R^n at the indices in $\mathbf{T} = [T_1, \ldots, T_q]$, denoted by $R^{\mathbf{T}}$, and Bob stores $R^{\mathbf{U}}$ for his indices $\mathbf{U} = [U_1, \ldots, U_q]$. We assume that they use a memory-efficient, implicit representation of the index set as described earlier for the private-key system, with on-line recomputation of the indices when necessary. In this way, Alice and Bob need approximately $\log q$ bits of memory each.

Because of the pairwise independent selection, both index sets can be determined from $2 \log n$ bits each. The descriptions of $\mathbf{T}$ and $\mathbf{U}$ exchanged on the public channel are therefore short. In order to apply Theorem 4 to the set $\{S_1, \ldots, S_l\} = \{T_1, \ldots, T_q\} \cap \{U_1, \ldots, U_q\}$ of common positions, we need the following lemma to make sure that also $S_1, \ldots, S_l$ have a uniform and pairwise independent distribution. It is easy to see that the expected number of common indices is $l = q^2/n$.

Lemma 5. *Let $T_1, \ldots, T_q$ and $U_1, \ldots, U_q$ be independent sequences of uniform and pairwise independent random variables, respectively, with alphabet $\{1, \ldots, n\}$ and distribution as described in Section 3, and let $S_1, \ldots, S_q$ be the sequence $T_1, \ldots, T_q$ restricted to those values occurring in $U_1, \ldots, U_q$, i.e. $S_j = T_j$ if there is an index h such that $U_h = T_j$ and $S_j = \omega$ otherwise. Then, the sequence $S_1, \ldots, S_q$ restricted to those positions different from ω is pairwise independent.*

Proof. Because the pairwise independence construction of Section 3 is used, no values in the $\mathbf{U}$ and $\mathbf{T}$ sequences are repeated. This implies

$$\mathrm{P}[T_i = x_1 \wedge T_j = x_2] \;=\; \frac{1}{n(n-1)}$$

for all $i, j \in \{1, \ldots, q\}$ and all $x_1, x_2 \in \{1, \ldots, n\}$ with $x_1 \neq x_2$. The sequence $S_1, \ldots, S_l$ satisfies

$$\mathrm{P}[S_i = x_1 \wedge S_j = x_2 | S_i \neq \omega \wedge S_j \neq \omega] \;=\; \frac{\mathrm{P}[S_i = x_1 \wedge S_j = x_2]}{\mathrm{P}[S_i \neq \omega \wedge S_j \neq \omega]} \tag{3}$$

for any $i, j \in \{1, \ldots, q\}$ and $x_1, x_2 \in \{1, \ldots, n\}$. Considering only those positions of the sequence $S_1, \ldots, S_q$ with values different from ω, we see that for any $i, j \in \{1, \ldots, q\}$ and all $x_1, x_2 \in$

$\{1, \ldots, n\}$ such that $x_1 \neq x_2$ and $S_i \neq \omega$ and $S_j \neq \omega$,

$$\begin{aligned} \mathrm{P}[S_i = x_1 \wedge S_j = x_2] &= \mathrm{P}[T_i = x_1 \wedge \exists h_1 : U_{h_1} = x_1 \wedge T_j = x_2 \wedge \exists h_2 : U_{h_2} = x_2] \\ &= \mathrm{P}[T_i = x_1 \wedge T_j = x_2] \cdot \mathrm{P}[\exists h_1, h_2 : U_{h_1} = x_1 \wedge U_{h_2} = x_2] \\ &= \frac{1}{n(n-1)} \cdot \frac{q(q-1)}{n(n-1)}. \end{aligned}$$

Furthermore, for all $i, j \in \{1, \ldots, q\}$, we have

$$\mathrm{P}[S_i \neq \omega \wedge S_j \neq \omega] \;=\; \mathrm{P}[\exists h_1, h_2 : U_{h_1} = T_i \wedge U_{h_2} = T_j] \;=\; \frac{q(q-1)}{n(n-1)}$$

because every pair of distinct x_1, x_2 occurs with the same probability in the sequence $U_1, \ldots, U_q$. Thus, the probability in (3) is equal to $\frac{1}{n(n-1)}$ for any $i, j \in \{1, \ldots, q\}$ and all $x_1 \neq x_2$, and the lemma follows. □

To illustrate a concrete example of the system, assume Alice and Bob both have access to a 40 Gbit/s broadcast channel. We need more network capacity for public key agreement than for private-key encryption to achieve a similar error probability. The channel is used for 2×10^5 seconds (about two days), thus $n = 8.6 \times 10^{15}$. Eve is allowed to store 1/2 PByte or $m = 4.5 \times 10^{15}$ bit. With $\Delta = 10^{-20}$ and error probabilities $\varepsilon_1 = 10^{-20}$ and $\varepsilon_2 = 10^{-3}$, the parameters are $\delta = 0.476$, $\rho = 0.077$, $l = 1.3 \times 10^7$, and $r = 5.0 \times 10^5$. In order to have l common indices on the average, Alice and Bob must store $q = \sqrt{ln} = 3.3 \times 10^{11}$ bit or about 39 GByte each (assuming the index sequences $\mathbf{T}$ and $\mathbf{U}$ are represented implicitly). The public communication between Alice and Bob consists of $2 \log n = 106$ bit in each direction for the selected indices plus 1.5 MByte in one direction for privacy amplification. Except with probability about 10^{-3}, Eve knows less than 10^{-20} bit about the 61 KByte secret key that Alice and Bob obtain.

Because l is on the order of the inverse squared error probability ε_2, the probabilities in the example are relatively large to keep the storage requirements of Alice and Bob at a reasonable level. Generating a shorter key does not help to reduce the storage space, which depends primarily on ε_2. It is an interesting open question whether Lemma 3 can be improved in order to reduce the influence on the error probability.

The large size of the hash function that has to be communicated for privacy amplification can be reduced by using "almost universal" hash functions based on almost k-wise independent random variables that can be constructed efficiently [1]. Such functions $g : \mathcal{X} \to \mathcal{Y}$ can be described with about $5 \log |\mathcal{Y}|$ instead of $\log |\mathcal{X}|$ bits.

7 Discussion

Our results show that unconditional security can be based on assumptions about the adversary's available memory. In essence, such a system exploits the capacity gap between fast communication and mass storage technology. We discuss a few implications of this fact.

First of all, generating random bits at a sufficiently high rate may be more expensive than merely transmitting them. However, a large investment in a random source can be amortized by the potentially high number of participants that can use the source simultaneously.

A drawback of our system is that the security margin is linear in the sense that memory costs are directly proportional to the offered storage capacity, at least up to technological advances. In most computationally secure encryption systems, the complexity of a brute-force attack grows exponentially in the length of the keys.

Our system is provably secure taking into account the current storage capacity of an adversary because the only possible attack is to store the broadcast data when it is sent. In contrast, most computationally secure systems can be broken retroactively, once better algorithms are discovered or faster processing becomes possible.

We have used the broadcast channel as an error-free black-box communication primitive in our system, although the legitimate users do not need its full functionality: They need not receive the complete broadcast, but only a small part of it. It is conceivable that a receiving device could be much simpler and less expensive if it can only synchronize and read a small, but arbitrary part of the traffic. Such receivers could also allow for a greater capacity of the channel.

The described protocols offer no resilience to errors on the broadcast channel. To take into account such errors, Alice and Bob can perform information reconciliation [4] on the selected subset. Methods for bounding the effect of this additional information provided to Eve are known [7].

The system rests on the gap between two technologies—fast communication and mass storage. Impressive future developments can be expected in both fields. We only mention the big potential of all-optical networks on one side and the recent developments in holographic and molecular storage on the other side.

Acknowledgment

It is a pleasure to thank Jan Camenisch, Markus Stadler, and Stefan Wolf for interesting discussions on this subject and the anonymous referees for helpful remarks.

References

[1] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, "Simple constructions of almost k-wise independent random variables," *Random Structures and Algorithms*, vol. 3, no. 3, pp. 289–304, 1992. Preliminary version presented at 31st FOCS (1990).

[2] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, pp. 1915–1923, Nov. 1995.

[3] G. Brassard and C. Crépeau, "25 years of quantum cryptography," *SIGACT News*, vol. 27, no. 3, pp. 13–24, 1996.

[4] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion," in *Advances in Cryptology — EUROCRYPT '93* (T. Helleseth, ed.), vol. 765 of *Lecture Notes in Computer Science*, pp. 410–423, Springer-Verlag, 1994.

[5] C. Cachin, *Entropy Measures and Unconditional Security in Cryptography.* Ph.D. dissertation No. 12187, ETH Zürich, 1997.

[6] C. Cachin, "Smooth entropy and Rényi entropy," in *Advances in Cryptology — EUROCRYPT '97* (W. Fumy, ed.), vol. 1233 of *Lecture Notes in Computer Science*, pp. 193–208, Springer-Verlag, 1997.

[7] C. Cachin and U. Maurer, "Linking information reconciliation and privacy amplification," *Journal of Cryptology*, vol. 10, no. 2, pp. 97–110, 1997.

[8] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 1991.

[10] C. Crépeau, "Efficient cryptographic protocols based on noisy channels," in *Advances in Cryptology — EUROCRYPT '97* (W. Fumy, ed.), vol. 1233 of *Lecture Notes in Computer Science*, pp. 306–317, Springer-Verlag, 1997.

[11] C. Crépeau and J. Kilian, "Achieving oblivious transfer using weakened security assumptions," in *Proc. 29th IEEE Symposium on Foundations of Computer Science (FOCS)*, 1989.

[12] R. Cruz, G. Hill, A. Kellner, R. Ramaswami, G. Sasaki, and Y. Yamabashi, Eds., "Special issue on optical networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, pp. 761–1052, June 1996.

[13] *Proc. 14th IEEE Symposium on Mass Storage Systems*, IEEE Computer Society Press, 1995.

[14] M. Luby, *Pseudorandomness and Cryptographic Applications.* Princeton University Press, 1996.

[15] M. Luby and A. Wigderson, "Pairwise independence and derandomization," Tech. Rep. 95-035, International Computer Science Institute (ICSI), Berkeley, 1995.

[16] J. L. Massey, "Contemporary cryptography: An introduction," in *Contemporary Cryptology: The Science of Information Integrity* (G. J. Simmons, ed.), ch. 1, pp. 1–39, IEEE Press, 1991.

[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. 1985 IEEE International Symposium on Information Theory*, p. 146, 1985.

[18] U. M. Maurer, "Conditionally-perfect secrecy and a provably-secure randomized cipher," *Journal of Cryptology*, vol. 5, pp. 53–66, 1992.

[19] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, pp. 733–742, 1993.

[20] C. J. Mitchell, "A storage complexity based analogue of Maurer key establishment using public channels," in *Cryptography and Coding: 5th IMA Conference, Cirencester, UK* (C. Boyd, ed.), vol. 1025 of *Lecture Notes in Computer Science*, pp. 84–93, Springer, 1995.

[21] N. Nisan, "Extracting randomness: How and why — a survey," in *Proc. 11th Annual IEEE Conference on Computational Complexity*, 1996.

[22] M. Rabin. Personal Communication, 1997.

[23] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, (Berkeley), pp. 547–561, Univ. of Calif. Press, 1961.

[24] R. L. Rivest, "Cryptography," in *Handbook of Theoretical Computer Science* (J. van Leeuwen, ed.), ch. 13, pp. 717–755, Elsevier, 1990.

[25] L. P. Seidman, "Satellites for wideband access," *IEEE Communications Magazine*, pp. 108–111, Oct. 1996.

[26] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, Oct. 1949.

[27] P. W. Shor, "Algorithms for quantum computation: Discrete log and factoring," in *Proc. 35th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 124–134, 1994.

[28] M. N. Wegman and J. L. Carter, "New hash functions and their use in authentication and set equality," *Journal of Computer and System Sciences*, vol. 22, pp. 265–279, 1981.

[29] D. Zuckerman, "Simulating BPP using a general weak random source," *Algorithmica*, vol. 16, pp. 367–391, 1996. Preliminary version presented at 32nd FOCS (1991).