Generalized Privacy Amplification*

Charles H. Bennett
IBM Research †

Gilles Brassard ‡
Université de Montréal §

Claude Crépeau
École Normale Supérieure ¶

Ueli M. Maurer, *Senior Member, IEEE* **
ETH Zürich ††

Abstract. This paper provides a general treatment of privacy amplification by public discussion, a concept introduced by Bennett, Brassard and Robert for a special scenario. Privacy amplification is a protocol that allows two parties to distill a secret key from a common random variable that is only partially secret, i.e. about which an eavesdropper has some partial information, where the two parties generally know nothing about the eavesdropper's information except that it satisfies a certain constraint. The results have applications to unconditionally-secure secret-key agreement protocols, quantum cryptography and to a generalized treatment of wire-tap and broadcast channels for a considerably strengthened definition of secrecy capacity.

Index terms. Cryptography, Secret-key agreement, Unconditional security, Privacy amplification, Wire-tap channel, Secrecy capacity, Rényi entropy, Universal hashing.

---

*The results of this paper were presented in part at the 1994 IEEE International Symposium on Information Theory, June 27 – July 1, 1994, Trondheim, Norway.

† IBM T.J. Watson Research Laboratory, Yorktown Heights, New York, NY 10598, USA. e-mail: bennetc@watson.ibm.com.

‡ Supported in part by NSERC's E. W. R. Steacie Memorial Fellowship and Québec's FCAR.

§ Département IRO, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal (Québec), Canada H3C 3J7. e-mail: brassard@iro.umontreal.ca.

¶ Laboratoire d'Informatique de l'École Normale Supérieure, C.N.R.S. URA 1327
45 rue d'Ulm, 75230 Paris CEDEX 05, France. e-mail: crepeau@dmi.ens.fr.

** Supported in part by the Swiss National Science Foundation (SNF).

†† Institute for Theoretical Computer Science, ETH Zürich, CH–8092 Zürich, Switzerland. e-mail: maurer@inf.ethz.ch.

I. Introduction

Privacy amplification is the art of distilling highly secret shared information (e.g. for use as a cryptographic key) from a larger body of shared information that is only partly secret. Let Alice and Bob be given a random variable W, e.g. a random n–bit string, while an eavesdropper, Eve, learns a correlated random variable V, providing at most $t < n$ bits of information about W, i.e. $H(W|V) \geq n - t$. The details of the eavesdropping distribution $P_{V|W}$, except that it satisfies this constraint as well as possibly some further constraints, are generally unknown to Alice and Bob. Alice and Bob wish to publicly choose a compression function $g : \{0,1\}^n \rightarrow \{0,1\}^r$ such that Eve's partial information on W and her complete information on g give her arbitrarily little information about $g(W)$. The size r of the secret that Alice and Bob can safely distill depends on the kind as well as the amount of information available to Eve. Assuming that W is a random n–bit string, various possible scenarios to consider are that Eve can obtain (1) t arbitrary bits of W, (2) t arbitrary parity checks of W, (3) the result of an arbitrary function mapping n–bit strings to t–bit strings, or (4) the string W transmitted through a binary symmetric channel with capacity t/n, i.e. with bit error probability ε satisfying $h(\varepsilon) = 1 - t/n$, where $h(\cdot)$ denotes the binary entropy function. We present a solution for a more general scenario of which all the above are special cases. In this scenario, Eve is allowed to specify an arbitrary distribution P_{VW} (unknown to Alice and Bob) subject to the only constraint that $R(W|V = v) \geq n - t$ with high probability (over values v), where $R(W|V = v)$ denotes the second-order conditional Rényi entropy [13, 26] of W given $V = v$ (see Section IV). More precisely, we show that for any $s > 0$ Alice and Bob can distill $n - t - s$ bits of secret key while keeping Eve's expected information about $g(W)$ smaller than $2^{-s}/\ln 2$ bits, where the expectation is computed over possible choices for the function g. This is quite remarkable because Alice and Bob do not know P_{VW}. It is important to point out that this result cannot be generalized to allowing Eve to obtain t arbitrary bits of information in Shannon's sense, i.e. $H(W|V) \geq n - t$, without further restriction, because in this case a possible strategy for Eve would be to receive the entire string W with probability t/n while receiving an independent random n–bit string with probability $1 - t/n$.

In the following, we provide the motivation and background for this research. One of the fundamental problems in cryptography is the generation of a shared secret key by two parties, Alice and Bob, not sharing a secret key initially. In large networks it is impractical and unrealistic to assume that a secure channel (such as a trusted courier) between Alice and Bob is available when the need for a secret key arises. Therefore we consider a scenario in which Alice and Bob are connected only over an insecure channel: all messages exchanged between Alice and Bob can be received completely by an eavesdropper Eve. However, it will be assumed throughout the paper that Eve cannot actively tamper with the channel by inserting or modifying messages, without being detected. Without such an assumption, secure communications, and in particular secret-key agreement, is impossible without an initially shared secret key. If the channel does not guarantee the authenticity of exchanged messages (for instance by speaker recognition on a telephone line) it can be guaranteed by well-known authentication techniques [32] that will not be discussed in this paper.

In this paper we are interested in *proving* the security of cryptographic schemes, in particular key agreement protocols. The significance of a proof of security of a cryptosystem depends crucially on three questions: (1) How strong is the definition of security? (2) How realistic are the assumptions about an eavesdropper's available information and computational resources? (3) How practical is the system? For most previous approaches to provable security, either the definition of security or the assumptions about an eavesdropper's information are not satisfactory, or the system is not practical. For instance, the implicit assumption in Wyner's wire-tap channel [33] that the channel from Alice to Eve is noisier than the channel from Alice to Bob is unrealistic in most applications, and the perfectly-secure one-time pad uses an impractically long key.

We will make no assumptions about the eavesdropper's computing power: we will consider unconditional or information-theoretic rather than computational security. There are two reasons for this: first, results obtained without any assumptions are stronger, and second, proving anything reasonable about the computational difficulty of breaking a cryptosystem appears to be completely out of reach of the current research in complexity theory and cryptography. Note that allowing Eve to use infinite computing power rules out public-key cryptography [15] as a possible technique for secret-key agreement. Of course, public-key cryptography is an important tool for implementing security on large networks, but no plausible restriction on Eve's computing power has been found that would allow a proof of security for any public key cryptosystem (or for any conventional system not already information-theoretically secure).

The strongest possible notion of security of a cryptosystem is *perfect secrecy* defined by Shannon [27]. A system is perfect if and only if the plaintext message M is statistically independent of the ciphertext C, i.e. $I(M;C) = 0$. Shannon proved the pessimistic result that perfect secrecy can be achieved only when the secret key K is at least as long as the plaintext message or, more precisely, when $H(K) \geq H(M)$. However, this proof is based on the assumption that an eavesdropper has available precisely the same information as the legitimate receiver, except for the secret key K, and this assumption is often unnecessarily restrictive.

For instance, when information is encoded in non-orthogonal states of a quantum system (such as non-orthogonal polarizations of a single photon), the uncertainty principle prevents an eavesdropper (or even the intended receiver) from extracting complete information from the quantum signal. This fact, and the related impossibility of making local measurements on entangled quantum states without disturbing their nonlocal correlations, can be exploited to limit an eavesdropper's information in key agreement protocols [3, 1, 7], as well as to prevent excessive indiscreet information flow in protocols for oblivious transfer [2] and bit commitment [8] between two cooperating but mutually suspicious parties. All these quantum cryptographic protocols, especially when implemented with real equipment [1, 25, 30], generate partly secret information that needs to be cleaned up before it can be fully exploited.

Similarly, Maurer [22] has suggested a key agreement protocol using a satellite to broadcast random bits that cannot be received completely reliably on the Earth, even with very expensive receiver technology. One can think of many other situations where the information seen by various parties is correlated without giving an eavesdropper perfect knowledge of all the other parties' information. Surprisingly, secret-key agreement does not require the correlation between Alice and Bob's information to be stronger than between Alice's and Eve's or Bob's and Eve's information.

Privacy amplification, and the key agreement it makes possible, have a broad range of cryptographic applications. This paper deals with key agreement, but of course an unconditionally-secure key-agreement protocol can be transformed into an unconditionally-secure encryption scheme by using the generated key as the key stream in the well-known one-time pad [27, 31]. A shared secret key is also required for unconditionally-secure authentication [32], which, like one-time-pad encryption, uses up key bits and renders them unfit for re-use. Finally, keys generated by privacy amplification serve as an intermediate tool in two-party protocols for oblivious transfer [2] and bit commitment [8]. One may wonder what purpose (or even what meaning) key agreement can have in the context of a two-party protocol, with no explicit eavesdropper. The answer is that in such protocols at least one of the parties (say Bob) is, if he is dishonest, trying to function simultaneously as a legitimate receiver (gaining authorized information) and as an eavesdropper (attempting to gain further unauthorized information), and the purpose of the protocol is to assist him in one role while defeating him in the other.

The paper is organized as follows. Unconditionally-secure key agreement is discussed in Section II, where the privacy amplification problem is motivated. In Section III, the general privacy amplification scenario considered in this paper is described and various types of side information available to an eavesdropper are discussed. Privacy amplification by universal hashing and the main results of this paper are presented in Section IV. In Section V, the gap between the achievable size of the secret key and the theoretical upper bound is illustrated, and in Section VI a technique for closing this gap is presented.

II. Unconditionally-secure secret-key agreement

Unconditionally-secure secret-key agreement by public discussion was introduced by Bennett, Brassard and Robert in [4, 5] and generalized by Maurer [22] who introduced a general information-theoretic model described below. It takes place in a scenario where Alice and Bob are connected by an insecure channel to which a passive eavesdropper Eve has perfect access, and where Alice, Bob and Eve know the correlated random variables X, Y and Z, respectively, which are distributed according to some joint probability distribution P_{XYZ}. The distribution P_{XYZ} may be partially under Eve's control. Quantum cryptography is an example of such a scenario where Eve's measurement influences the outcome of the random experiment. In contrast, in the satellite scenario discussed in [22], where X, Y and Z result from a binary random string broadcast by a satellite and received by Alice, Bob and Eve over individual and partially independent noisy channels, Eve has no control over P_{XYZ}. A surprising fact demonstrated in [22] is that even if Eve's channel is much superior compared to Alice's and Bob's channels, unconditionally-secure secret-key agreement is possible.

A key agreement protocol for such a scenario generally consists of three phases. The first phase, introduced in [22] and called advantage distillation in [10], is needed when neither Alice nor Bob has an advantage compared to Eve, i.e. if neither $I(X;Y) > I(X;Z)$ nor $I(X,Y) > I(Y;Z)$. After this phase, involving a sequence of messages summarized in a random variable C, Alice can compute a string W from X and C about which Bob has less uncertainty than Eve: $H(W|XC) = 0$ and $H(W|YC) < H(W|ZC)$.

In the second phase, often referred to as information reconciliation [1, 9], Alice and Bob exchange redundant information and apply error-correction techniques in order for Bob to be able to learn W with very high probability but such that Eve is left only with incomplete information about it. Generally, Alice can send a bit string D whose length L is slightly larger than $H(W|YC)$, and Eve's remaining uncertainty is hence at least $H(W|ZCD) \geq H(W|ZC) - L$, which is generally positive.

In the third phase, which is called privacy amplification and is the subject of this paper, Alice and Bob distill from W a shorter string K about which Eve has only a negligible amount of information. For instance, Alice and Bob publicly agree on a function g that becomes known to Eve, and they let $K = g(W)$. Whether or not Alice and Bob can generate such a secret key depends on the particular information about W known to Eve and, to only a surprisingly small extent, on Alice's and Bob's knowledge about the type of Eve's information.

The key agreement scenario we describe should be contrasted with previously proposed scenarios for unconditionally-secure message transmission. Wyner [33] considered a communications scenario in which Alice can send information to Bob over a discrete memoryless channel such that a wire-tapper Eve can receive Bob's channel output only through an additional cascaded independent channel, giving Eve a degraded version of what Bob receives. Wyner's model and results were generalized by Csiszár and Körner [14] who considered a discrete memoryless broadcast channel for which Eve's received message is not necessarily a degraded version of the legitimate receiver's message, but Bob's channel must be better than Eve's. In these scenarios, as in most of those considered in this paper, perfect secrecy cannot be achieved exactly but arbitrarily closely, i.e. Eve is left with a positive (albeit arbitrarily small) amount of information.

For characterizing "almost-perfect" secrecy, Wyner [33], and Csiszár and Körner [14] defined the secrecy capacity of a broadcast channel scenario as the maximal rate at which Alice can send secret information to Bob while keeping the rate at which Eve obtains this information arbitrarily small, and they characterized the secrecy capacity for general broadcast channels.

However, the results of [33] and [14] are not in the most desirable form from a practical viewpoint for several reasons:

- $\diamond$ They depend on non-constructive existence proofs similar to that of Shannon's noisy coding theorem.
- $\diamond$ The definition of secrecy capacity is not strong enough: because the rate rather than Eve's absolute amount of information is bounded, and especially in view of the fact that a very large block length must be used, it is conceivable that, although the rate at which Eve receives information is small, all or part of the information she is really interested in is contained in this small fraction of the whole message.
- $\diamond$ A model in which exactly the same random experiment is repeated many times is generally good enough for a scenario where the goal is reliable communication. However, in a cryptographic scenario involving an intelligent opponent, a more general treatment allowing the opponent to choose from a large variety of strategies is much more desirable and useful.
- $\diamond$ Finally, the assumption that an eavesdropper (initially) has less information than Alice and Bob is unrealistic in many cases.

The techniques presented in this paper allow the first problem to be solved when the main channel from Alice to Bob is noiseless. While the privacy amplification technique is constructive and efficient, achieving the secrecy capacity in a general scenario still involves non-constructive existence arguments for the information reconciliation phase. This paper also allows to solve the second and third problem: Eve's *total* expected amount of information, rather than the rate, is proved to be arbitrarily small, and our results can handle very general scenarios of eavesdropping information. The last problem can be solved by using advantage-distillation techniques discussed in [22]. The solution of the first problem is also considered in [20].

III. Privacy amplification by public discussion

Let W and V be two random variables with joint distribution P_{VW}, where W (which takes on values in the set $\mathcal{W}$) is known to Alice and Bob and V summarizes all Eve's information about W. Eve might be able to choose which partial information about W she would like to see, i.e. P_{VW} could partially be under Eve's control. In other words, Eve might be able to choose P_{VW} from a set of admissible such distributions, where P_W could or could not be the same for all possible choices. If P_W cannot be influenced by Eve, her choice can be characterized by a conditional distribution $P_{V|W}$. Alice and Bob generally do not know P_{VW} and possibly not even P_W, but they know that P_{VW} satisfies a certain constraint. For instance, in quantum cryptography this constraint follows from the uncertainty principle of quantum physics.

Alice and Bob publicly agree on a function $g : \mathcal{W} \rightarrow \{0,1\}^r$ for a suitable r and compute the r–bit secret key $K = g(W)$. In general, g is selected randomly from an appropriate set $\mathcal{G}$ of functions in order to avoid that Eve knows g before deciding about her strategy for accessing information about W. In other words, the compression function is actually a random variable G taking on as values functions $\mathcal{W} \rightarrow \{0,1\}^r$ from $\mathcal{G}$. Rather than statements for specific functions g, our results will be statements in the form of averages over choices of G or, alternatively, about the probability of picking a function g with a certain property. Surprisingly, Alice and Bob need not know anything about P_{VW} or P_W other than the fact that P_{VW} satisfies the assumed constraint.

We are interested in upper bounds of the form $I(K;GV) \leq \varepsilon$ for some arbitrarily small ε, provided P_{VW} satisfies a given constraint. More precisely, we will derive bounds of the form

$$H(K|G, V = v) \geq r - \varepsilon \tag{1}$$

for a very small ε, which hold for the concrete value v of V known to Eve, rather than only on the average, provided that $P_{W|V=v}$ satisfies a given constraint. Results of this form are therefore stronger and generally imply average results if the constraint on $P_{W|V=v}$ is satisfied for all v or at least with high probability. Note that (1) implies that given all Eve's information, K is virtually uniformly distributed. If (1) is satisfied for a set of values v with total probability at least $1 - \delta$ for a small δ, then K has almost maximal entropy: $(1 - \delta)r - \varepsilon \leq H(K) \leq r$.

The length r of the secret key K that can be distilled by Alice and Bob of course depends on P_{VW}. More generally, it depends on the type of constraint that P_{VW} must satisfy. The more strongly W and V are correlated, the smaller is r. Similarly, the more restrictive the constraint on Eve's strategy for selecting P_{VW}, the larger is r in general. In the following we discuss various types of constraints on P_{VW}. In order to have a common denominator for these examples, we let W be a random n–bit string which takes on all 2^n possible values with equal probability, and we allow Eve to obtain t bits of information about W. However, our results apply to general distributions P_{VW}. The main motivation for this discussion is to point out that all previous results only applied to constraints that are not satisfied in a realistic scenario. This gap is closed in this paper.

The extreme cases are not interesting: when Eve cannot obtain any information about W (i.e. $t = 0$) then Alice and Bob can let $K = W$ and hence $r = n$. On the other hand, if Eve knows W precisely ($t = n$) then it is not surprising that Alice and Bob cannot generate a secret key. (Nevertheless, the proof of this fact and that $r > n$ is impossible even when $t = 0$ are not completely trivial [22].)

If Eve knows (i.e. V consists of) at most t physical bits of W and Alice and Bob know which bits Eve knows, then they can use the remaining $n - t$ bits as the secret key. However, if Eve can choose the positions of the t bits secretly, then the problem is more interesting. This situation has been considered independently in [4] and [12]. Provided $t < n$, it is always possible for Alice and Bob to distill a shorter string $K = g(W)$ about which it is guaranteed that Eve has no information whatsoever (except for its length). The function g may even be chosen in full view of Eve *before* she decides which t bits to access. However, K must in general be much shorter than W even for small values of t (i.e. $r \ll n - t$). For instance, it is proven in [12] that when $t = 2$ the length of K can be at most roughly $2/3$ of that of W. Larger values of t have been investigated by

Friedman [16] and Stinson[28]. It is known that the most efficient functions g must be nonlinear for some values of n and t [29].

A less restrictive constraint on P_{VW} is that Eve is allowed to obtain t arbitrary parities of bits of W. A still much less restrictive constraint on P_{VW}, which includes the former example, is that Eve is allowed to secretly specify an arbitrary function $e : \{0,1\}^n \to \{0,1\}^t$ of her choice and to receive $e(W)$. In other words, Eve can obtain t arbitrary bits of deterministic information about W. This case was solved in 1985 [4] (see [5] for more details).

However, this result cannot be applied in many realistic scenarios because in general Eve's information is probabilistic rather than deterministic. For instance, in a scenario that arises in the secret-key agreement protocols discussed in [22], Eve can receive the bits of W through a binary symmetric channel with fixed bit error probability ε. Similarly, in more general scenarios arising in quantum cryptography, Eve can receive the bits of W through binary symmetric channels whose bit error probability she can control, subject to a global constraint over the bit error probabilities for all the n bits. Our results apply to these as well as more general scenarios.

In the process of further generalizing the constraint on the type of the t bits of information available to Eve such that Alice and Bob can still generate an almost secret string K of length close to $n - t$, it is clear that we cannot allow Eve to obtain t arbitrary bits of information in the most general sense of Shannon's information theory. For instance, if P_W is assumed to be fixed, we cannot allow her to specify an arbitrary distribution $P_{V|W}$ subject to the constraint $I(W;V) \leq t$. This is too much freedom for Eve because an admissible strategy for her would be to obtain an n–bit string V that is equal to W with probability t/n but an independent random string with probability $1 - t/n$, and where Eve does not learn which of the two cases has occurred. It is easy to see that this "channel" gives Eve slightly less than t bits of information about W. However, it is straightforward to compute that Eve will have more than $rt/n - 1$ bits of information about K, regardless of which function g is publicly chosen to transform the n–bit string W into an r–bit string K, because $g(V)$ is equal to K with probability at least t/n. In other words, Eve will have essentially the same proportion of information about K as she had about W.

This leads to the following natural question: Which is the most general information measure $\tilde{I}$ such that when Eve is allowed to obtain t arbitrary bits of information about W according to this measure, i.e. to specify an arbitrary distribution $P_{V|W}$ subject to the constraint $\tilde{I}(W;V) \leq t$, then Alice and Bob can nevertheless distill essentially $n - t$ bits of secret key. A very general measure, which is based on Rényi entropy of order 2 [26], is discussed below.

IV. Universal hashing and Rényi entropy

It was first discovered by Bennett, Brassard and Robert [5] that an important technique for privacy amplification against deterministic eavesdropping is universal hashing, a concept introduced by Carter and Wegman [11]. Our method also draws on the work of Impagliazzo, Levin, Luby and Zuckerman [18, 19], who used Rényi entropy to quantify the randomness produced by universal hashing, but in the context of quasirandom number generation rather than privacy amplification.

Definition 1. A class $\mathcal{G}$ of functions $\mathcal{A} \to \mathcal{B}$ is *universal*$_2$ ("universal" for short) if, for any distinct x_1 and x_2 in $\mathcal{A}$, the probability that $g(x_1) = g(x_2)$ is at most $1/|\mathcal{B}|$ when g is chosen at random from $\mathcal{G}$ according to the uniform distribution.

For $r = n$, the class consisting only of the identity function is trivially a universal class, although there are others. For the more interesting case $r < n$, the class of all functions from $\{0,1\}^n$ to $\{0,1\}^r$ is universal, but it is not useful because there are too many functions in that class (it takes $r2^n$ bits to specify a function). A more useful universal class is that of all *linear* functions from $\{0,1\}^n$ to $\{0,1\}^r$ [11]. These functions can be described by $r \times n$ matrices M over $GF(2)$, i.e. by rn bits. Other universal classes, which are more economical in terms of the number of bits needed to specify them, are discussed in [11, 32]. Such a class, requiring only n bits to specify a function, is given in the following lemma.

Lemma 1. *Let a be an element of $GF(2^n)$ and also interpret x as an element of $GF(2^n)$. Consider the function $\{0,1\}^n \to \{0,1\}^r$ assigning to an argument x the first r bits of the element ax of $GF(2^n)$. The class of such functions for $a \in GF(2^n)$ is a universal class of functions for $1 \le r \le n$.*

Proof. For any two $x_1, x_2 \in GF(2^n)$ with $x_1 \neq x_2$, and for any nonzero $a \in GF(2^n)$ we have $[ax_1]_r = [ax_2]_r$ if and only if $[a(x_1 - x_2)]_r$ is all zero, where $[y]_r$ for $y \in GF(2^n)$ denotes the first r bits of y. For each of the 2^{n-r} y with $[y]_r$ all zero, there is exactly one a with $a(x_1 - x_2) = y$. Hence, for any $x_1 \neq x_2$ there are exactly 2^{n-r} values a for which $[ax_1]_r = [ax_2]_r$. The probability of choosing such an a is $(2^{n-r})/2^n = 2^{-r}$. $\square$

Definition 2. Let X be a random variable with alphabet $\mathcal{X}$ and distribution P_X. The *collision probability* $P_c(X)$ of X is defined as the probability that X takes on the same value twice in two independent experiments:

$$P_c(X) = \sum_{x \in X} P_X(x)^2.$$

The *Rényi entropy of order two* of X [13, 26], henceforth simply called Rényi entropy for short, is defined as the negative logarithm of the collision probability of X:

$$R(X) = -\log_2 P_c(X).$$

For an event $\mathcal{E}$, the collision probability and the Rényi entropy of X conditioned on $\mathcal{E}$, $P_c(X|\mathcal{E})$ and $R(X|\mathcal{E})$, are defined naturally as the collision probability and the Rényi entropy, respectively, of the conditional distribution $P_{X|\mathcal{E}}$. The Rényi entropy conditioned on a random variable, $R(X|Y)$, is the expected value of the conditional Rényi entropy:

$$R(X|Y) = \sum_y P_Y(y)\, R(X|Y=y).$$

In order to contrast Rényi entropy with the standard entropy measure defined by Shannon, we will refer to the latter as "Shannon entropy" throughout the paper. Note that Rényi entropy (like Shannon entropy) is always positive. $R(X)$ can equivalently be expressed as $R(X) = -\log_2 E[P_X(X)]$, where $E[\cdot]$ denotes the expected value. Shannon entropy $H(X)$ can be expressed similarly as $H(X) = -E[\log_2 P_X(X)]$. It follows from Jensen's inequality (see [6, p. 428]) that Rényi entropy is upper bounded by the Shannon entropy.

Lemma 2. *For every discrete probability distribution P_X,*

$$R(X) \leq H(X),$$

with equality if and only if P_X is the uniform distribution over $\mathcal{X}$ or a subset of $\mathcal{X}$. Moreover, for every distribution P_{XY},

$$R(X|Y) \leq H(X|Y).$$

At first it seems natural to extend the analogy between Rényi and Shannon entropies to the notion of information. In other words, it is tempting to define the mutual Rényi information between X and Y to be $I_R(X;Y) = R(X) - R(X|Y)$. However, this notion is not symmetric as $R(X) - R(X|Y)$ is different from $R(Y) - R(Y|X)$ in general. Nevertheless, it makes sense to say that $R(X) - R(X|Y)$ represents the amount of Rényi information about X that is provided by Y. In particular, we say that the eavesdropper "has t bits of Rényi information about W" if W is uniformly distributed over $\{0,1\}^n$ and if her information V is such that $R(W|V) = n - t$.

The following theorem demonstrates that Rényi entropy plays the role of a general information measure that we are looking for.

Theorem 3. *Let X be a random variable over the alphabet $\mathcal{X}$ with probability distribution P_X and Rényi entropy $R(X)$, let G be the random variable corresponding to the random choice (with uniform distribution) of a member of a universal class of hash functions $\mathcal{X} \to \{0,1\}^r$, and let $T = G(X)$. Then*

$$H(T|G) \;\geq\; R(T|G) \;\geq\; r - \log_2(1 + 2^{r-R(X)}) \;\geq\; r - \frac{2^{r-R(X)}}{\ln 2}.$$

Note that G is a random variable and that the quantity $H(T|G) = H(G(X)|G)$ is an average over all choices of the function g. It is possible that even when $R(X) \gg r$, $H(G(X)|G = g) = H(g(X))$ differs from r by a non-negligible amount for some g, but such a g can occur only with negligible probability.

Proof. The first inequality follows from Lemma 2. The second inequality can be proved as follows:

$$\begin{aligned} R(G(X)|G) &= \sum_g P_G(g) R(G(X)|G = g) \\ &= \sum_g P_G(g)(-\log_2 P_c(G(X)|G = g)) \\ &\geq -\log_2 \left(\sum_g P_G(g) P_c(G(X)|G = g) \right) \end{aligned}$$

where the last step follows from Jensen's inequality. The sum in the last term is equal to the probability that $g(x_1) = g(x_2)$ if g is chosen at random according to P_G and x_1 and x_2 are chosen at random, independently of each other and of g, according to P_X. Therefore we have

$$\begin{aligned} \sum_g P_G(g) P_c(G(X)|G = g) &= \text{Prob}[G(X_1) = G(X_2)] \\ &= \text{Prob}[X_1 = X_2] + \\ &\quad\ \text{Prob}[X_1 \neq X_2] \cdot \text{Prob}[G(X_1) = G(X_2) \mid X_1 \neq X_2] \\ &\leq P_c(X) + (1 - P_c(X)) \cdot 2^{-r} \\ &< 2^{-R(X)} + 2^{-r} \\ &= 2^{-r}\left(1 + 2^{r-R(X)}\right). \end{aligned} \tag{2}$$

Here the first inequality follows from the fact that the class of functions is universal and by noting that $1/|\mathcal{B}| = 2^{-r}$ according to Definition 1. The second and third inequalities of the theorem now follow immediately by taking logarithms on both sides of (2) and from the inequality $\log_2(1 + y) \leq y/\ln 2$, respectively. □

This theorem clearly applies also to conditional probability distributions such as $P_{W|V=v}$ discussed above. We therefore have the following result on privacy amplification.

Corollary 4. *Let P_{VW} be an arbitrary probability distribution. If Eve's Rényi entropy $R(W|V = v)$ about W is known to be at least h and Alice and Bob choose $K = G(W)$ as their secret key, where G is chosen at random from a universal class of hash functions from $\mathcal{W}$ to $\{0, 1\}^r$, then*

$$H(K|G, V = v) \ \geq\ r - \log_2(1 + 2^{r-h}) \ \geq\ r - \frac{2^{r-h}}{\ln 2}.$$

Thus we see that Eve's entropy of the secret key K is close to maximal, i.e. her distribution of K is close to uniform. Alternatively we could say that her information about K, namely $H(K) - H(K|G, V = v)$, is arbitrarily small. More precisely, if $r < h$, then Eve's total information about K decreases exponentially in the excess compression $h - r$. It should be pointed out that Corollary 4 cannot be generalized to Rényi entropy conditioned on a random variable, i.e. both

$$H(K|GV) \ \geq \ r - \log_2 \left(1 + 2^{r-R(W|V)}\right)$$

as well as the weaker inequality

$$H(K|GV) \ \geq \ r - \frac{2^{r-R(W|V)}}{\ln 2}$$

are false in general. However, if V takes on a value v satisfying $R(W|V = v) \geq h$ with probability at least $1 - \delta$, then we have $H(K|GV) \geq (1 - \delta)(r - \log_2(1 + 2^{r-R(W|V)}))$.

The following is a (slightly strengthened) reformulation of Theorem 10 of [5] but the proof given here is much simpler than the original proof.

Corollary 5. *Let W be a random n-bit string with uniform distribution over $\{0,1\}^n$, let $V = e(W)$ for an arbitrary eavesdropping function $e : \{0,1\}^n \rightarrow \{0,1\}^t$ for some $t < n$, let $s < n - t$ be a positive safety parameter, and let $r = n - t - s$. If Alice and Bob choose $K = G(W)$ as their secret key, where G is chosen at random from a universal class of hash functions from $\{0,1\}^n$ to $\{0,1\}^r$, then Eve's expected information about the secret key K, given G and V, satisfies*

$$I(K; GV) \ \leq \ 2^{-s}/\ln 2.$$

Note that, in contrast to Corollary 4, this result is (and must be) stated as an average over the values of V. Note also that Alice's and Bob's strategy does not depend on e and hence privacy amplification works even if they have no information about e, provided they know an upper bound on t.

Proof. For $v \in \{0,1\}^t$, let c_v be the number of $w \in \{0,1\}^n$ that are consistent with v, i.e. satisfying $e(w) = v$. Given $V = v$, all consistent w are equally likely candidates for W and hence $P_{W|V=v} = 1/c_v$ for all w with $e(w) = v$. Therefore we have $P_c(W|V = v) = c_v \cdot (1/c_v)^2 = 1/c_v$ and $R(W|V = v) = \log_2 c_v$ and thus, according to Corollary 4,

$$H(K|G, V = v) \ \geq \ r - \frac{2^{r - \log_2 c_v}}{\ln 2} \ = \ r - \frac{2^r}{c_v \cdot \ln 2} \ .$$

Averaging over values of v and using $P_V(v) = c_v 2^{-n}$ we obtain

$$\begin{aligned} I(K;GV) &= H(K) - H(K|GV) \\ &\leq r - \sum_{v\in\{0,1\}^t} P_V(v) H(K|G, V = v) \\ &\leq \sum_{v\in\{0,1\}^t} c_v 2^{-n} \frac{2^r}{c_v \cdot \ln 2} \\ &= 2^{-n+t+r} / \ln 2 \ = \ 2^{-s} / \ln 2. \end{aligned}$$

□

V. The gap between Rényi and Shannon entropy

Let us now investigate the implication of Theorem 3 in a genuine case of probabilistic information. Assume that Alice and Bob share a random n–bit string W (uniformly distributed over the n-bit strings) and that Eve can receive the output V when W is sent through a binary symmetric channel with bit error probability ε. Hence we have

$$P_{W|V=v}(w) = (1-\varepsilon)^{n-d(v,w)} \varepsilon^{d(v,w)},$$

where $d(v,w)$ is the Hamming distance between v and w. It is easy to check that Rényi entropy, like Shannon entropy, is additive for independent random variables. It follows that

$$H(W|V = v) = nh(\varepsilon) = -n(\varepsilon \log_2 \varepsilon + (1-\varepsilon)\log_2(1-\varepsilon))$$

and

$$R(W|V = v) = -n \log_2((1-\varepsilon)^2 + \varepsilon^2)$$

for all v.

Example: Consider an example relevant to quantum cryptography [1]: $\varepsilon = \sin^2 22\frac{1}{2}^\circ \approx 0.15$. In this case, $R(W) - R(W|V = v) \approx 0.585\,n$ whereas $H(W) - H(W|V = v) \approx 0.399\,n$ for all v. Note that these bounds also apply to the averages over choices for v: $I(W;V) \approx 0.399\,n$ and $R(W) - R(W|V) \approx 0.585\,n$. Observe that this eavesdropping strategy gives Eve significantly more Rényi information than it gives her Shannon information.

It is instructive to contrast the described eavesdropping scenario with the following eavesdropping strategy. Assume Eve had the opportunity to obtain $\tilde{V}$ consisting of $n/2$ arbitrary bits of W of her choice. This would give her $R(W) - R(W|\tilde{V} = \tilde{v}) = 0.5\,n$ bits of Rényi information and also $H(W) - H(W|\tilde{V} = \tilde{v}) = 0.5\,n$ bits of Shannon information for all $\tilde{v} \in \{0,1\}^{n/2}$. Note that in these examples we have $H(W|V = v) = H(W|V)$ and $R(W|V = v) = R(W|V)$ for all v, and $H(W|\tilde{V} = \tilde{v}) = R(W|\tilde{V} = \tilde{v}) = H(W|\tilde{V}) = R(W|\tilde{V})$ for all $\tilde{v}$. Therefore

$$R(W) - R(W|V) \ > \ R(W) - R(W|\tilde{V})$$

whereas

$$H(W) - H(W|V) \; < \; H(W) - H(W|\tilde{V}).$$

In other words, the second strategy gives Eve more Shannon but less Rényi information than the first strategy. This leads to the following natural question: if the eavesdropper could choose between the two strategies, which one should she adopt? This choice may seem somewhat arbitrary but in fact it is very similar to a real choice facing the eavesdropper on the quantum channel [1].

If Eve knew that Alice and Bob are going to use W directly as key for a one-time pad encipherment, she would probably prefer the deterministic strategy since it would give her more information about the cleartext, especially if she knows which bits of the cleartext will be most interesting. But the situation is less clear if Alice and Bob intend to apply privacy amplification to W in order to distill a shorter key K. Indeed, Theorem 3 suggests that it is the Rényi information available to the eavesdropper on the original string W, not the Shannon information, that determines the size of K that allows Alice and Bob to reach a predetermined secrecy level: the length of K can be at most $R(W|V=v) - s$ bits in order for Eve's expected information to be less than $2^{-s}/\ln 2$ bits.

If the first eavesdropping strategy is adopted, then the resulting secret K can be of length at most $0.415\,n - s$. On the other hand, as many as $0.5n - s$ highly secret bits can be distilled from W if Alice and Bob know that Eve used the deterministic strategy. Thus it appears that even though the first strategy discloses fewer bits of Shannon information to Eve concerning W, it gives her bits that are more resistant to privacy amplification, with the net result that it forces Alice and Bob to sacrifice more bits by universal hashing in order to produce a highly secret key.

Wyner's celebrated wire-tap channel [33] provides an alternative (albeit non-constructive and less secure) approach to privacy amplification when Eve uses the first strategy. Wyner's technique allows Alice and Bob to derive a secret key K at a rate arbitrarily close to $h(\varepsilon)$ in the limit as $n \to \infty$ while giving Eve only a negligible *fraction* of information about K. When $\varepsilon = \sin^2 22\frac{1}{2}^\circ$, this means that K can be as long as $0.6\,n$ bits for sufficiently large n, which is better than the length achievable if Eve used the deterministic strategy. Therefore, the conclusion reached in the previous paragraph was incorrect: Eve's better strategy is the one that gave her more Shannon information about W, regardless of the amount of Rényi information involved. What is wrong with Theorem 3? There are two possible explanations for this phenomenon.

- ◇ Privacy amplification by universal hashing is not optimal when Eve's information is obtained through a binary symmetric channel.
- ◇ Privacy amplification by universal hashing is fine, but the *analysis* provided by Theorem 3 is overly pessimistic: even though shrinking the string as much as stipulated by the theorem is guaranteed to leave the eavesdropper with a very small amount of expected information, less shrinkage would be sufficient to achieve the desired secrecy level.

We shall see in the next section that the second explanation is correct: privacy amplification by universal hashing is essentially optimal for a binary symmetric channel scenario. More generally, it is optimal for any scenario in which a fixed random experiment is repeated many times independently [10, 23]. The length of the secret key that can be distilled is roughly equal to Eve's Shannon entropy about W.

VI. Spoiling knowledge

Although Rényi entropy is a useful tool, reducing the amount of such information available to the eavesdropper is not a goal in itself: it is the amount of expected *Shannon* information that we really wish to make arbitrarily small. By virtue of Lemma 2 we know that this goal can be attained by reducing Rényi information, but this may be overkill. To illustrate this fact, consider a random W chosen with uniform distribution over the n–bit strings, and let Eve's distribution $P_{W|V=v}$ over the n–bit strings be defined by

$$P_{W|V=v}(w) = \begin{cases} 2^{-n/4} & \text{if } w = v \\ \dfrac{1-2^{-n/4}}{2^n-1} & \text{otherwise.} \end{cases}$$

Although $P_c(W|V=v) > (2^{-n/4})^2 = 2^{-n/2}$ and hence $R(W|V=v) < n/2$, which means that Eve has more than half the Rényi information about W, she has essentially no Shannon information: it is straightforward to check that $H(W|V=v) > n(1-2^{n/4})$ and hence Eve has an exponentially small amount of information about W. Therefore privacy amplification is completely unnecessary: Alice and Bob may safely use W directly as their secret!

This example also illustrates a counter-intuitive property of Rényi entropy, which we are going to exploit in the sequel. Unlike Shannon entropy, the average Rényi entropy can increase when side information is obtained. In other words,

$$R(X|Y) > R(X)$$

is possible. In the previous example, consider an oracle who gives Eve for free the random variable U defined by $U=0$ if $w=v$ and $U=1$ otherwise. With probability $2^{-n/4}$, this gives Eve complete information about W, and her Rényi information jumps from roughly $n/2$ bits to n bits. However, with overwhelming complementary probability $1-2^{-n/4}$, Eve becomes almost completely ignorant about W: all she knows is that W is different from V. More formally, we have $R(W|U=0, V=v) = 0$ and $R(W|U=1, V=v) = \log_2(2^n-1)$ and hence

$$\begin{aligned} R(W|U, V=v) &= (1-2^{-n/4})\log_2(2^n-1) \\ &\geq (1-2^{-n/4})(n-2^{1-n}) \\ &> n - n2^{-n/4} - 2^{1-n} \end{aligned}$$

provided $n \geq 1$. Thus learning the value of U has the effect for Eve of decreasing her Rényi information from about $n/2$ bits to an expected exponentially small amount no bigger than $n2^{-n/4} + 2^{1-n}$ bits.

Such information, which increases Eve's Rényi entropy, will be called *spoiling knowledge* because Eve would rather *not* know it if she cares to maximize her Rényi information. We use it as a conceptual step in the analysis of privacy amplification to allow Alice and Bob to distill a secret key K from a given string W that is potentially much longer than suggested by considering $R(W|V=v)$ and Corollary 4.

Spoiling knowledge U about a random variable X can be used effectively if the possible outcomes u of the spoiling variable can be partitioned into two sets: a low-probability set small enough to be neglected, and a high-probability set consisting of u outcomes that give little Rényi information about X; and separately estimating the contributions of the two sets to the after-hashing Shannon entropy $H(G(X)|G)$. This leads to the following generalization of Theorem 3.

Theorem 6. *Let X, G and T be defined as in Theorem 3, and let U be another random variable, jointly distributed with X according to an arbitrary distribution P_{XU} for which the marginal distribution for X coincides with P_X. Then*

$$H(T|G) \;\geq\; r - \sum_u P_U(u) \log_2 \left(1 + 2^{r-R(X|U=u)}\right)$$

and

$$H(T|G) \;\geq\; r - \sum_u P_U(u) \cdot \min\left\{r, \frac{2^{r-R(X|U=u)}}{\ln 2}\right\}.$$

In particular, $H(T|G)$ is lower bounded by the maximum, over choices of P_{XU} consistent with the given P_X, of either expression on the right-hand side.

Proof. The first inequality follows from

$$H(T|G) \;\geq\; H(T|GU) \;=\; \sum_u P_U(u) \cdot H(T|G, U=u)$$

and by application of Theorem 3 to lower bound the terms $H(T|G, U=u)$ by $r - \log_2(1 + 2^{r-R(X|U=u)})$. The second inequality follows from the trivial bound $H(T|G, U=u) \geq 0$ and from the weaker inequality of Theorem 3, namely $H(T|G, U=u) \geq r - \frac{2^{r-R(X|U=u)}}{\ln 2}$. □

This raises the following interesting question, which is suggested as an open problem: which side information U maximizes the right-hand side of Theorem 6 for a given distribution P_X? It appears helpful if possible to design U such that for every $U=u$, the distribution $P_{X|U=u}$ is uniform or close to uniform.

This theorem clearly applies also to conditional probability distributions such as $P_{W|V=v}$, which arise when privacy amplification is called upon because the eavesdropper has obtained information on the string common to Alice and Bob.

Corollary 7. *Let W, V, G and K be as defined in Corollary 4 and let U be another random variable, jointly distributed with W and V according to some distribution P_{UVW} for which the marginal distribution of $[V, W]$ coincides with P_{VW}. Then*

$$H(K|G, V=v) \;\geq\; r - \sum_u P_{U|V}(u, v) \cdot \log_2 \left(1 + 2^{r-R(W|U=u,V=v)}\right)$$

and

$$H(K|GV) \;\geq\; r - \sum_{u,v} P_{UV}(u,v) \cdot \log_2\left(1 + 2^{r - R(W|U=u,V=v)}\right).$$

In particular, $H(K|G, V = v)$ and $H(K|GV)$ are lower bounded by the maximum of the respective right-hand sides over choices of P_{UVW} consistent with the given P_{VW}.

The bound on $H(K|GV)$ can be very close to the maximum r if $R(W|U = u, V = v)$ is large (compared to r) with very high probability (over choices of u and v).

The concept of spoiling knowledge can now be applied to perform a more careful analysis of eavesdropping through a binary symmetric channel; a technique introduced in [2]. Remember that

$$R(W|V = v) \;=\; -n \log_2((1-\varepsilon)^2 + \varepsilon^2)$$

for all v. Consider the spoiling knowledge $U = d(W, v)$ consisting of the Hamming distance between W and the particular value v known to Eve. Given $U = u$, all $\binom{n}{u}$ strings w at distance u from v are equally likely candidates for W, i.e.

$$R(W|U = u, V = v) \;=\; \log_2 \binom{n}{u}.$$

For any λ between 0 and 1, we have

$$\binom{n}{\lambda n} \geq \frac{2^{nh(\lambda)}}{\sqrt{8n\lambda(1-\lambda)}} \geq \frac{2^{nh(\lambda)}}{\sqrt{2n}}. \tag{3}$$

The left inequality is proved as Lemma 7 in [21, p. 309] and the second inequality follows from $8\lambda(1-\lambda) \leq 2$. Consider now an arbitrarily small positive constant $\delta \leq \varepsilon$. By the law of large numbers we have

$$(\varepsilon - \delta)n < d(W, v) < (\varepsilon + \delta)n,$$

and hence also

$$\binom{n}{u} > \binom{n}{(\varepsilon - \delta)n},$$

except with probability exponentially small in n. Hence inequality (3) implies that

$$\binom{n}{u} > \frac{2^{nh(\varepsilon-\delta)}}{\sqrt{2n}},$$

and hence that

$$R(W|U = u, V = v) \;>\; nh(\varepsilon - \delta) - \log_2 \sqrt{2n},$$

except with probability exponentially small in n. Consider now any fixed $\gamma > 0$ and let $\delta > 0$ be so that $h(\varepsilon - \delta) > h(\varepsilon) - \gamma$. We conclude that

$$R(W|U = u, V = v) \;>\; (h(\varepsilon) - \gamma)n$$

for all sufficiently large n, except with probability exponentially small in n. This exponentially small probability cannot contribute more than an exponentially small amount of expected Shannon information for Eve. The theorem below follows immediately.

Theorem 8. *For all positive ε and γ, there exists a positive α such that if Alice and Bob share a random n–bit string W, which Eve receives through a binary symmetric channel with bit error probability ε, and if they apply privacy amplification with universal hashing to obtain an r–bit string K where $r = \lfloor (h(\varepsilon) - \gamma) n \rfloor$, then for all sufficiently large n, Eve's expected information about K is at most $2^{-\alpha n}$.*

This theorem answers the question posed in Section V: if Eve wishes to minimize the amount of key Alice and Bob can safely agree on through privacy amplification, it is better for her to obtain $n/2$ bits of her choice from Alice's and Bob's n–bit string x, rather than to access each bit of this string through a binary symmetric channel with $\varepsilon = \sin^2 22\frac{1}{2}^\circ$. On the other hand, if W itself were a winning lottery number, or some other information that would be useful to Eve only if she guessed it exactly, she should use the latter strategy.

VII. Conclusions

The results of this paper have several applications in cryptography, namely in all scenarios where an eavesdropper Eve is known for some reason to have only incomplete information about a certain random variable. In general, Alice and Bob will not have complete information about the random variable either, and they will first need to generate a common string W about which Eve has only partial information. A general model of correlated random variables available to Alice, Bob and Eve is discussed in [22]. Using privacy amplification by public discussion as described in this paper, they can subsequently distill a secret key about which Eve has arbitrarily little information.

A typical scenario of the type described is that arising in quantum cryptography [1], where the uncertainty principle of quantum physics prevents Eve (and also Bob) from obtaining complete information about the polarization of a photon sent by Alice. Another scenario, discussed in [22], is one in which Alice, Bob and Eve can all receive a random sequence broadcast by a satellite over partially independent channels, where Eve's channel could be much more reliable than the other two channels. We hope that our results on privacy amplification will open new unexpected approaches to practical unconditional security in cryptography.

Acknowledgements

The term "spoiling knowledge" was coined by Silvio Micali in the context of zero-knowledge proofs. We are also grateful to Christian Cachin, Ivan Damgård, Martin Gander, Jean–Marc Robert, Louis Salvail, Miklós Sántha, and William Wootters for stimulating discussions on privacy amplification.

References

[1] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail and J. Smolin, "Experimental quantum cryptography", *Journal of Cryptology*, Vol. 5, no. 1, 1992, pp. 3–28.

[2] C. H. Bennett, G. Brassard, C. Crépeau and M.–H. Skubiszewska, "Practical quantum oblivious transfer", *Advances in Cryptology — Proceedings of Crypto '91*, Lecture Notes in Computer Science, Vol. 576, Springer–Verlag, Berlin, 1992, pp. 351–366.

[3] C. H. Bennett, G. Brassard and N. D. Mermin, "Quantum cryptography without Bell's theorem", *Physical Review Letters*, Vol. 68, 1992, pp. 557–559.

[4] C. H. Bennett, G. Brassard and J.–M. Robert, "How to reduce your enemy's information", *Advances in Cryptology — Proceedings of Crypto '85*, August 1985, Lecture Notes in Computer Science, Vol. 218, Springer–Verlag, Berlin, 1986, pp. 468–476.

[5] C. H. Bennett, G. Brassard and J.–M. Robert, "Privacy amplification by public discussion", *SIAM Journal on Computing*, Vol. 17, no. 2, April 1988, pp. 210–229.

[6] R. E. Blahut, *Principles and Practice of Information Theory*, Reading, MA: Addison-Wesley, 1987.

[7] G. Brassard, "Cryptology column — Quantum cryptography: A bibliography", *Sigact News*, Vol. 24, no. 3, 1993, pp. 16–20. [An expanded version is available from the author.]

[8] G. Brassard, C. Crépeau, R. Jozsa and D. Langlois, "A quantum bit commitment scheme provably unbreakable by both parties," *Proceedings of 34th IEEE Symposium on Foundations of Computer Science*, 1993, pp. 362–371.

[9] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion", *Advances in Cryptology — Proceedings of Eurocrypt '93*, 1993, Lecture Notes in Computer Science, Vol. 765, Springer Verlag, Berlin, 1994, pp. 410–423.

[10] C. Cachin and U. M. Maurer, "Linking information reconciliation and privacy amplification", to appear in *Advances in Cryptology - EUROCRYPT '94*, Lecture Notes in Computer Science, Springer-Verlag, Berlin.

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

[12] B. Chor, O. Goldreich, J. Håstad, J. Friedman, S. Rudich and R. Smolensky, "The bit extraction problem or t–resilient functions", *Proceedings of 26th IEEE Symposium on Foundations of Computer Science*, October 1985, pp. 396–407.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., 1981.

[14] I. Csiszár and J. Körner, "Broadcast channels with confidential messages", *IEEE Transactions on Information Theory*, Vol. IT–24, no. 3, 1978, pp. 339–348.

[15] W. Diffie and M.E. Hellman, "New directions in cryptography", *IEEE Transactions on Information Theory*, Vol. IT–22, 1976, pp. 644–654.

[16] J. Friedman, "On the bit extraction problem", *Proceedings of 33rd IEEE Symposium on Foundations of Computer Science*, October 1992, pp. 314–319.

[17] M. Gander and U. M. Maurer, "On the secret-key rate of binary random variables (extended abstract)", *Proceedings of 1994 IEEE International Symposium on Information Theory*, Trondheim, Norway, June 27–July 1, 1994, p. 351.

[18] R. Impagliazzo, L. A. Levin and M. Luby, "Pseudo-random generation from one-way functions", *Proceedings of 21st Symposium on Theory of Computer Science*, 1989, pp. 12–24.

[19] R. Impagliazzo and D. Zuckerman, "How to recycle random bits", *Proceedings of 30th IEEE Symposium on Foundations of Computer Science*, October 1989, pp. 248–253.

[20] V. Korzhik and V. Yakovlev, "Nonasymptotic estimates of information protection efficiency for the wire-tap channel concept", *Advances in Cryptology — Proceedings of Auscrypt '92*, December 1992, Lecture Notes in Computer Science, Vol. 718, Springer–Verlag, Berlin, 1993, pp. 185–195.

[21] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North–Holland, New York, 1977.

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

[23] U. M. Maurer "The strong secret key rate of discrete random triples", *Communications and Cryptography, Two Sides of one Tapestry*, R.E. Blahut et al. (editors), Kluwer Academic Publishers, 1994, pp. 271–285.

[24] R. J. McEliece, "The theory of information and coding", *Encyclopedia of Mathematics and its Applications*, Vol. 3, Addison-Wesley, Reading, MA, 1977.

[25] A. Muller, J. Breguet and N. Gisin, "Experimental demonstration of quantum cryptography using polarized photons in optical fibre over more than 1 km" *Europhysics Letters*, Vol. 23, no. 6, 20 August 1993, pp. 383–388.

[26] A. Rényi, "On measures of entropy and information", *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 1961, pp. 547–561.

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

[28] J. Bierbrauer, K. Gopalakrishnan, and D. R. Stinson, "Orthogonal arrays, resilient functions, error correcting codes and linear programming bounds", submitted to SIAM J. on Discrete Mathematics.

[29] D. R. Stinson and J. L. Massey, "An infinite class of counterexamples to a conjecture concerning non-linear resilient functions", *Journal of Cryptology*, to appear.

[30] P. D. Townsend, "Secure key distribution system based on quantum cryptography", *Electronics Letters*, Vol. 30, no. 10, 12 May 1994, pp. 809–811.

[31] G. S. Vernam, "Cipher printing telegraph systems for secret wire and radio telegraphic communications", *Journal of the American Institute of Electrical Engineers*, Vol. 55, 1926, pp. 109–115.

[32] 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, 1981, pp. 265–279.

[33] A. D. Wyner, "The wire-tap channel", *Bell System Technical Journal*, Vol. 54, no. 8, 1975, pp. 1355–1387.