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 ††

May 31, 1995

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 process that allows two parties to distill a secret key from a common random variable about which an eavesdropper has partial information. 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 and quantum cryptography, and they yield results on 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, Quantum cryptography.

---

*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, perhaps for use as a cryptographic key, from a larger body of shared information that is only partially secret. Let Alice and Bob be given a random variable W, such as 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 distribution P_{VW} are generally unknown to Alice and Bob, except that it satisfies this constraint as well as possibly some further constraints. They may or may not know P_W. 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 $K = g(W)$, except with negligible probability (over possible choices for g). The resulting K is virtually uniformly distributed given all Eve's information; it can hence be used safely as a cryptographic key.

The size r of the secret that Alice and Bob can 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 bit error probability ε satisfying $h(\varepsilon) = 1 - t/n$, and hence with capacity 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 [15], [27] of W, given $V = v$ (see Section IV). For any $s < n - t$, Alice and Bob can distill $r = n - t - s$ bits of secret key $K = G(W)$ while keeping Eve's information about K exponentially small in s, by publicly choosing the compression function G (which is now a random variable) at random from a suitable class of maps into $\{0,1\}^{n-t-s}$. More precisely, we show that $H(K|G, V = v) \geq r - 2^{-s}/\ln 2$, provided only that $R(W|V = v) \geq n - t$. It is shown that this result cannot be generalized to allowing Eve to obtain t arbitrary bits of information in Shannon's sense without further restriction: $H(W|V = v) \geq n - t$ is not a sufficient restriction on P_{VW} for privacy amplification to be possible.

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. In large networks it is impractical and unrealistic to assume that a secure channel (such as a trusted courier) is available between Alice and Bob 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. The validity of this assumption can be guaranteed by well-known unconditionally-secure authentication techniques [32] that will not be discussed here, provided Alice and Bob initially share a short unconditionally-secure secret key. Assuming the existence of such a short secret key for the purpose of authentication does not render key-agreement protocols useless: such protocols can be interpreted as allowing an arbitrary unconditionally-secure expansion of a short key. Other means such as speaker recognition on a telephone line could also be used in practice.

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 perfectly-secure one-time pad uses an impractically long key. Moreover, assumptions of the type that the channel from Alice to Bob is less noisy than the channel from Alice to Eve [16], [33] is unrealistic in many applications.

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 computational complexity theory and cryptography. Note that allowing Eve to use unlimited computing power rules out public-key cryptography [17] 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.

The strongest possible notion of security of a cryptosystem is *perfect secrecy* defined by Shannon [28]. 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 M 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 [4], [2], [9], as well as to prevent excessive indiscreet information flow in protocols for oblivious transfer [3] and bit commitment [10] between two cooperating but mutually suspicious parties. All these quantum cryptographic protocols, especially when implemented with real equipment [2], [26], [30], [23], generate partly secret information that needs to be cleaned up (e.g., by privacy amplification) before it can be fully exploited.

Similarly, Maurer [24] 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 [28], [31]. Also, privacy amplification plays a major role in various two-party protocols for oblivious transfer [3] and bit commitment [10]. Finally, a shared secret key is required for unconditionally-secure authentication [32], which, like one-time-pad encryption, uses up key bits and renders them unfit for re-use. Thus we see that unconditionally-secure authentication and privacy amplification can benefit from one another: the former serves the latter to make sure that the public discussion between Alice and Bob is not tampered with by Eve, and some of the resulting secret shared key bits can be used for subsequent authentication purposes. A short initial secret key between Alice and Bob is necessary to prime this system and avoid the apparent vicious circle.

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 [5], [6] and generalized by Ahlswede and Csiszár [1] and by Maurer [24] 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. A surprising fact demonstrated in [24] 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, provided Alice and Bob know P_{XYZ}.

A key agreement protocol for such a scenario generally consists of three phases. The first phase, introduced in [24] and called advantage distillation in [12], is needed when neither Alice nor Bob has an advantage compared to Eve, i.e. if neither $I(X;Y) > I(X;Z)$ nor $I(X,Y) > I(Y;Z)$ or, more precisely, if the forward key-capacity stated in Theorem 1 of [1] is zero. 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 [5], [2], [11], 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 $H(W|YC)$ so that $H(W|YCD) \approx 0$. Eve's remaining uncertainty will be $H(W|ZCD) \geq H(W|ZC) - L$, which can be substantially 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 [16] 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 less noisy than Eve's. In these scenarios, as in most of those considered in this paper, perfect secrecy cannot be achieved exactly.

For characterizing "almost-perfect" secrecy, Wyner [33] and Csiszár and Körner [16] 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 [16] are not in the most desirable form for several reasons:

- ◇ The assumption that the legitimate users' channel is less noisy than an eavesdropper's channel [16] is unrealistic in many cases.
- ◇ The results of [33] and [16] depend on random coding arguments. While the step of selecting such a random code is efficient, decoding it is highly inefficient by all known techniques.
- ◇ 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 more desirable and useful.
- ◇ Finally, 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 when the secrecy capacity should closely be achieved, 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.

The first problem can be solved by using advantage-distillation techniques discussed in [24]. The techniques presented in this paper allow the second problem to be solved when the main channel from Alice to Bob is noiseless because privacy amplification is entirely practical and efficient. The solution of this problem is also considered in [21]. The problem of achieving secrecy capacity efficiently when information reconciliation and perhaps also advantage distillation are necessary is currently under investigation; see [11] for preliminary findings. The third problem is addressed in that our results can handle quite general scenarios of eavesdropping information.

Addressing the last problem, the results of this paper were shown in [25] to imply that in the key agreement scenarios of [1] and [24], a stronger definition of key-capacity and secret-key rate, respectively, can be used instead of the old definition, by requiring Eve's total amount of information to be arbitrarily small (a notion pioneered in [5]). For the case of broadcast channels with confidential messages [16], in which the goal is to transmit securely a meaningful message rather than merely to agree on a secret key, one can use the broadcast channel together with privacy amplification to generate secret key bits at a rate equal to the secrecy capacity [25]. Because the broadcast channel can also be used as a regular channel from Alice to Bob (with capacity at least equal to the secrecy capacity), the shared secret key can be used as the key in a one-time pad encryption of the message to be transmitted. Note that because Eve's distribution is known, it is possible to fix a suitable compression function beforehand, at the same time when the code for the broadcast channel is fixed. Let the *strong* secrecy capacity of a broadcast channel be defined like the secrecy capacity [16], except that the wire-tapper's total information—rather than the rate—is required to be arbitrarily small. The above arguments show that the strong secrecy capacity is at least half the secrecy capacity, but it is an open problem whether they are equal.

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 $g : \mathcal{W} \to \{0,1\}^r$ 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 $\mathcal{W} \to \{0,1\}^r$ 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.

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 for Eve: $(1-\delta)(r-\varepsilon) \leq H(K|GV) \leq r$.

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 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 an information-theoretically secret key. (Nevertheless, the proofs of this fact and that $r > n$ is impossible even when $t = 0$ are not completely trivial [24].)

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 [5] and [14]. 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 [14] that when $t = 2$ the length of K can be at most roughly $2/3$ of that of W. Larger values of t have also been investigated [18], [7]. 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} 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 [5] (see [6] 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 [24], Eve can receive the bits of W through a binary symmetric channel with fixed bit error probability ε. Similarly, in 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 W is assumed to be uniformly distributed over n–bit strings, we cannot allow her to specify an arbitrary distribution $P_{V|W}$ subject only to the constraint $I(W;V) \leq t$, nor even to the stronger constraint $H(W|V = v) \geq n - t$ for the specific value v that she gets to observe. 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 n–bit random string otherwise, 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 greater than 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_{VW} 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 general measure, which is based on Rényi entropy of order 2 [27], is discussed below.

IV. Universal hashing and Rényi entropy

It was first discovered by Bennett, Brassard and Robert [6] that an important technique for privacy amplification against deterministic eavesdropping is universal hashing, a concept introduced by Carter and Wegman [13]. Our method also draws on the subsequent work of Impagliazzo, Levin, Luby and Zuckerman [19], [20], 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 [13]. 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.

When $\mathcal{A} = \mathcal{B}$, the class consisting only of the identity function is trivially a universal class, although there are others. In all cases, the class of all functions from $\mathcal{A}$ to $\mathcal{B}$ is universal, but it is not useful because there are too many functions in that class. A more useful universal class is that of all *linear* functions from $\{0,1\}^n$ to $\{0,1\}^r$ [13]. 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 [13], [32]. Such a class, requiring only n bits to specify a function, is given in the following lemma whose proof is ommitted.

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 all such functions for $a \in GF(2^n)$ is a universal class of functions for $1 \leq r \leq n$.*

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:

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

The *Rényi entropy of order two* ("Rényi entropy" for short) of X [15], [27] 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 [8, p. 428]) that Rényi entropy is upper bounded by the Shannon entropy, a result already known to Rényi.

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. Moreover, $R(X) - R(X|Y)$ can be negative, as we shall see in Section VI.

The following theorem demonstrates that Rényi entropy can play 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 $\mathcal{X} \rightarrow \{0,1\}^r$, and let $Q = G(X)$. Then*

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

Note that G is a random variable and the quantity $H(Q|G) = H(G(X)|G)$ 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 other two inequalities are proved as follows:

$$\begin{aligned} R(G(X)|G) &= \sum_g P_G(g) R(G(X)|G = g) \\ &= \sum_g P_G(g)(-\log_2 P_c(G(X)|G = g)) \\ &\geq -\log_2 \left(\sum_g P_G(g) P_c(G(X)|G = g) \right) \end{aligned} \tag{2}$$

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

$$\begin{aligned} \sum_g P_G(g) P_c(G(X)|G = g) &= \mathrm{Prob}[G(X_1) = G(X_2)] \\ &= \mathrm{Prob}[X_1 = X_2] + \\ &\quad\ \mathrm{Prob}[X_1 \neq X_2] \cdot \mathrm{Prob}[G(X_1) = G(X_2) \mid X_1 \neq X_2] \\ &\leq P_c(X) + (1 - P_c(X)) \cdot 2^{-r} \\ &< 2^{-R(X)} + 2^{-r} \\ &= 2^{-r} \left(1 + 2^{r - R(X)}\right). \end{aligned} \tag{3}$$

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 from (2) by taking logarithms on both sides of (3), 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 and let v be a particular value of V observed by Eve. If Eve's Rényi entropy $R(W|V = v)$ about W is known to be at least c 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*

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

Thus we see that when $r < c$, Eve's entropy of the secret key K is close to maximal, i.e. her distribution of K is close to uniform. In particular, her information about K, namely $H(K) - H(K|G, V = v)$, is arbitrarily small. More precisely, her total information about K decreases exponentially in the excess compression $c - r$. 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 the probability is at least $1 - \delta$ that V takes on a value v satisfying $R(W|V = v) \geq c$, then we have

$$H(K|GV) \ \geq \ (1 - \delta)(r - \log_2(1 + 2^{r-c})).$$

and therefore

$$I(K; GV) \ \leq \ \delta r + (1 - \delta) \log_2(1 + 2^{r-c}) \leq \delta r + 2^{r-c}/ \ln 2.$$

The following is a slightly strengthened version of Theorem 10 of [6]. 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 $e : \{0,1\}^n \to \{0,1\}^t$ 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 Corollary 4 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 [2]:
$\varepsilon = \sin^2 22\frac{1}{2}^\circ \approx 0.15$. 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: $R(W) - R(W|V) \approx 0.585\,n$ and $H(W) - H(W|V) \approx 0.399\,n$. Observe that this eavesdropping strategy reduces Eve's Rényi entropy significantly more than it does her Shannon entropy.

It is instructive to contrast the described eavesdropping scenario with the following eavesdropping strategy, also relevant to quantum cryptography. If Eve could obtain $\tilde{V}$ consisting of $n/2$ arbitrary bits of W of her choice, this would reduce her Rényi entropy by $R(W) - R(W|\tilde{V} = \tilde{v}) = 0.5\,n$ bits and would also reduce her Shannon entropy by $H(W) - H(W|\tilde{V} = \tilde{v}) = 0.5\,n$ 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, Eve obtains a better reduction in Rényi entropy with the first strategy, but she obtains a better reduction in Shannon entropy with the second strategy. It is hence an interesting question which strategy is better for Eve.

If the first eavesdropping strategy is adopted, then it seems that the resulting secret K can be of length at most $0.415\,n - s$, where s is Alice's and Bob's safety parameter. 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 second strategy. However, Wyner's results on the wire-tap channel [33] provide an alternative (albeit inefficient 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. This apparent contradiction is resolved in the following section.

VI. Auxiliary random variables

Our goal is to leave Eve with negligible Shannon information about the secret key. By virtue of Lemma 2 we know that this can be accomplished by making her Rényi entropy close to maximal, but this may be overkill. To illustrate this, consider a random variable 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 $R(W|V=v) < n/2$ is far from maximal, 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 $K = W$ can be used directly as secret key, with no need to sacrifice more than half the key length to privacy amplification, as Corollary 4 would suggest.

This example also illustrates a counter-intuitive property of Rényi entropy, which we are going to exploit in the sequel. Unlike Shannon entropy, Rényi entropy can increase when it is conditioned on a random variable. 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 entropy falls from roughly $n/2$ bits to 0 bits: $R(W|U=0, V=v) = 0$. However, with overwhelming complementary probability $1-2^{-n/4}$, we have $R(W|U=1,V=v)=\log_2(2^n-1)$ 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 $n \geq 1$, which is approximately twice as large as the unconditioned Rényi entropy $R(W|V=v)$.

The fact that conditioning on an auxiliary random variable can increase Rényi entropy can be used to prove that privacy amplification allows 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.

Consider the scenario and notation of Theorem 3. An auxiliary random variable U is useful if its range can be partitioned into two sets: a set of very small total probability and its complement consisting of values u for which the Rényi entropy of X conditioned on $U=u$ is high. This leads to the following generalization of Theorem 3.

Corollary 6. *Let X, G and Q 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*

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

and

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

In particular, $H(Q|G)$ 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

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

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

This raises the following interesting question, which is suggested as an open problem: which auxiliary random variable U maximizes the right-hand side of Corollary 6 for a given distribution P_X? Because this auxiliary random variable denotes information that would increase Eve's Rényi entropy of X with high probability, and hence make her more vulnerable to privacy amplification via Theorem 3, it has been suggested to call it *spoiling knowledge*, a term coined by Silvio Micali in the context of zero-knowledge proofs.

Corollary 6 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 defined as 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). These results generalize a technique introduced in [3] to perform a more careful analysis of eavesdropping through a binary symmetric channel with error probability ε. Recall that

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

for all v. Consider the auxiliary random variable $U=d(W,v)$ 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

$$\binom{n}{\lambda n} \geq \frac{2^{nh(\lambda)}}{\sqrt{8n\lambda(1-\lambda)}} \geq \frac{2^{nh(\lambda)}}{\sqrt{2n}}. \tag{4}$$

The left inequality is proved as Lemma 7 in [22, 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 (4) 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 $h(\varepsilon - \delta) > h(\varepsilon) - \gamma/3$. We conclude that

$$R(W|U = u, V = v) \; > \; (h(\varepsilon) - \gamma/2)n$$

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 Shannon information for Eve. The next result 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 $2^{-\alpha n}$ bits.*

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 about which Eve has only partial information. A general model of correlated random variables available to Alice, Bob and Eve is discussed in [1] and [24]. 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 [25].

A typical scenario of the type described is that arising in quantum cryptography [2], 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 [24], 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.

A further application of privacy amplification is for achieving higher secrecy in the broadcast channel scenario of [16]. Whenever the secrecy capacity is positive, so is the strong secrecy capacity for which the wire-tapper's total amount (rather than the rate) of information about the exchanged message is required to be arbitrarily small.

Acknowledgements

We would like to thank Christian Cachin, Ivan Damgård, Martin Gander, Jean–Marc Robert, Louis Salvail, Miklós Sántha and William Wootters for stimulating discussions on privacy amplification, and an anonymous referee for many helpful comments.

References

[1] R. Ahlswede and I. Csiszár, "Common Randomness in information theory and cryptography – part I: secret sharing", *IEEE Transactions on Information Theory*, Vol. IT–39, 1993, pp. 1121 – 1132.

[2] 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.

[3] 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.

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

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

[6] 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.

[7] J. Bierbrauer, K. Gopalakrishnan, and D. R. Stinson, "Orthogonal arrays, resilient functions, error correcting codes and linear programming bounds", submitted to *SIAM Journal on Discrete Mathematics*, 1994.

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

[9] G. Brassard, "Cryptology column — Quantum cryptography: A bibliography", *Sigact News*, Vol. 24, no. 3, 1993, pp. 16 – 20.

[10] 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.

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

[12] C. Cachin and U. M. Maurer, "Linking information reconciliation and privacy amplification", *Journal of Cryptology*, to appear.

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

[14] 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.

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

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

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

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

[19] 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.

[20] 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.

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

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

[23] C. Marand and P. D. Townsend, "Quantum key distribution over distances up to 30 km", submitted to *Optics Letters*, 1995.

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

[25] 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.

[26] 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.

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

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

[29] D. R. Stinson and J. L. Massey, “An infinite class of counterexamples to a conjecture concerning non-linear resilient functions”, *Journal of Cryptology*, Vol. 8, no. 3, 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 – 810.

[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.