Preprint, 30-Oct-95. To appear in *Journal of Cryptology.*

Linking Information Reconciliation and Privacy Amplification*

Christian Cachin and Ueli M. Maurer

Institute for Theoretical Computer Science
ETH Zürich
CH-8092 Zürich, Switzerland
E-mail: {cachin,maurer}@inf.ethz.ch

Abstract

Information reconciliation allows two parties knowing correlated random variables, such as a noisy version of the partner's random bit string, to agree on a shared string. Privacy amplification allows two parties sharing a partially secret string about which an opponent has some partial information, to distill a shorter but almost completely secret key by communicating only over an insecure channel, as long as an upper bound on the opponent's knowledge about the string is known. The relation between these two techniques has not been well understood. In particular, it is important to understand the effect of side-information, obtained by the opponent through an initial reconciliation step, on the size of the secret key that can be distilled safely by subsequent privacy amplification. The purpose of this paper is to provide the missing link between these techniques by presenting bounds on the reduction of the Rényi entropy of a random variable induced by side-information. We show that, except with negligible probability, each bit of side information reduces the size of the key that can be safely distilled by at most two bits. Moreover, in the important special case of side information and raw key data generated by many independent repetitions of a random experiment, each bit of side information reduces the size of the secret key by only about one bit. The results have applications in unconditionally secure key agreement protocols and in quantum cryptography.

1 Introduction

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 the presence of an adversary Eve. One generally assumes that Eve can eavesdrop on the communication between Alice and Bob who are connected only by a public channel. It is easy to see that if this public channel is not assumed to be authenticated, then such key agreement is impossible. We therefore assume that any modification or insertion of messages can be detected by Alice and Bob.

This problem can be solved by applying public-key cryptography [9], where one assumes that Eve's computing power is limited and that certain problems are computationally difficult. In the recent years, key agreement protocols have been developed that are secure against adversaries with unlimited computing power [1, 10]. The motivation for investigating such protocols is two-fold: First, one avoids having to worry about the generality of a particular computational model, which is of some concern in view of the potential realizability of quantum computers [5, 13]. Second, no strong rigorous results on the difficulty of breaking a cryptosystem have

*This research was supported by the Swiss National Science Foundation. A preliminary version of this paper was presented at Eurocrypt '94, May 9–12, Perugia, Italy.

been proved, and this problem continues to be among the most difficult ones in complexity theory.

Unconditionally secure secret-key agreement [10, 11] takes place in a scenario where Alice, Bob and Eve know the correlated random variables X, Y and Z, respectively, distributed according to some joint probability distribution that may be under partial control of Eve (as for instance in quantum cryptography [1]). One possible scenario considered by Maurer [10] is that X, Y and Z result from a binary random string broadcast by a satellite and received by Alice, Bob and Eve over noisy channels. Secret-key agreement is possible even when Eve's channel is much more reliable than Alice's and Bob's channels.

A key agreement protocol for such a scenario generally consists of three phases:

Advantage Distillation: [10] The purpose of the first phase is to create a random variable W about which both Alice or Bob have more information than Eve. Advantage distillation is only needed when such a W is not immediately available from X and Y, for instance, when Eve's channel is superior in the above satellite scenario. Alice and Bob create W by exchanging messages, summarized as the random variable C, over the public channel.

Information Reconciliation [1, 6]: To agree on a string T with very high probability, Alice and Bob exchange redundant error-correction information U, such as a sequence of parity checks. After this phase, Eve's (incomplete) information about T consists of Z, C and U.

Privacy Amplification [2, 3]: In the final phase, Alice and Bob agree publicly on a compression function G to distill from T a shorter string S about which Eve has only a negligible amount of information. Therefore, S can subsequently be used as a secret key.

Information reconciliation and privacy amplification are fundamental for unconditionally secure key agreement and quantum key distribution. Advantage distillation has not yet been used in quantum cryptography because the considered scenarios assume that Alice and Bob have an advantage compared to Eve.

If after the first phase Alice knows a string about which Bob has more information than Eve, Alice and Bob can choose W to be this string. In other words, using information-theoretic terms, W is a random variable such that $H(W|XC) = 0$ and $H(W|YC) < H(W|ZC)$. In such a case, Bob tries to determine W from Y and the reconciliation string U. (Note that $H(U) \geq H(W|YC)$ is a necessary condition.) Hence reconciliation serves to establish $H(W|YCU) \approx 0$ while Eve still has a substantial amount of uncertainty about W: $H(W|ZCU) > 0$. After privacy amplification, $H(S)$ should be as large as possible, and Eve's information about S should be arbitrarily close to zero: $I(S; ZCUG) = H(S) - H(S|ZCUG) \approx 0$. Note that Alice and Bob can both compute S, i.e., $H(S|WG) = 0$.

In the following, let $V = [Z, C]$ summarize Eve's total knowledge about W before reconciliation. For deriving lower bounds on Eve's final information about the secret key S one can either consider a particular value $V = v$ that Eve knows or one can average over all possible values of V. Note that results for a particular $V = v$, which will be considered in this paper, are stronger than averaging results because they are known to hold for the very instance of the protocol execution. In other words, Eve's information about W is modeled by the probability distribution $P_{W|V=v}$ about which Alice and Bob have some incomplete knowledge. In particular, they know a lower bound on the Rényi entropy (see below) of the distribution $P_{W|V=v}$ with high probability but they do not know v.

It is known [2] that the Rényi entropy after reconciliation with $U = u$ (i.e., of the distribution $P_{W|V=v,U=u}$) is a lower bound on the size of the secret key that can be distilled safely by privacy amplification. This paper is concerned with understanding the reduction of the Rényi entropy induced by the side information U, either for a particular value $U = u$, or averaged over all values of U. Although this question is fundamental for any proof in the area of key agreement protocols, it has previously not been well understood because the behavior of Rényi entropy is different from that of Shannon entropy with respect to side-information. Existing proofs such as the ingenious Big-Brother argument of [1] work only for particular probability distributions and reconciliation protocols.

The paper is organized as follows. Section 2 reviews privacy amplification and the definition of Rényi entropy. Section 3 presents upper bounds on the reduction of Rényi entropy due to side-information for arbitrary probability distributions. Non-interactive reconciliation protocols with uniform and close-to-uniform probability distributions are investigated in Section 4. These results are applied in Section 5 to analyze the important class of scenarios in which a given random experiment is repeated many times independently.

2 Review of Privacy Amplification and Rényi Entropy

We assume that the reader is familiar with the notion of entropy and the basic concepts of Shannon's information theory [4, 8]. In privacy amplification, a different entropy measure, *Rényi entropy*, is of central importance [2]. To distinguish Rényi entropy from entropy in the sense of Shannon, we will always refer to the latter as *Shannon entropy*. All logarithms in this paper are to the base 2, and entropies are thus measured in bits.

Privacy amplification was introduced by Bennett, Brassard and Robert [3] and investigated further in [2], and can be described as follows. Assume Alice and Bob share an n-bit string W about which an eavesdropper Eve has incomplete information characterized by a probability distribution $P_{W|V=v}$ over the n-bit strings, where v denotes the particular value taken on by the random variable V summarizing her side-information. For instance, Eve might have received some bits or parities of bits of W, she might have eavesdropped on some of the bits of W through a binary symmetric channel, or have some more complicated type of information about W. Alice and Bob have some knowledge of the distribution $P_{W|V=v}$, but they do not know exactly what is compromised about their string. Using a public channel, which is totally susceptible to eavesdropping but immune to tampering, they wish to agree on a function $g : \{0,1\}^n \rightarrow \{0,1\}^r$ such that Eve, despite her partial knowledge about W and complete knowledge of g, almost certainly knows nearly nothing about $g(W)$. This process transforms a partially secret n-bit string W into a highly secret but shorter r-bit string $g(W)$ which can be used as a secret key.

The method for selecting the function g proposed in [3] is to choose it at random from a publicly-known *universal class of hash functions* mapping n-bit strings to r-bit strings. Universal hash functions were introduced by Carter and Wegman [7]. A class G of functions $\mathcal{A} \rightarrow \mathcal{B}$ is called *universal* 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 with uniform distribution from G.

Bennett, Brassard, Crépeau and Maurer [2] showed that the Rényi entropy (defined below) of Eve's distribution about W provides a lower bound on the size r of the secret key distillable from W by privacy amplification with a universal hash function.

Definition 1. 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* (or "Rényi entropy" for short) of X [12, 2] is defined as the negative logarithm of the collision probability of X:

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

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

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

Equivalently, $R(X)$ can be expressed as $R(X) = -\log E[P_X(X)]$, where $E[\cdot]$ denotes the expected value. Shannon entropy $H(X)$ can be expressed similarly as $H(X) = -E[\log P_X(X)]$. It follows from Jensen's inequality (see [8]) that Rényi entropy is upper bounded by the Shannon entropy, a fact known to Rényi:

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

with equality if and only if P_X is the uniform distribution over $\mathcal{X}$ or a subset of $\mathcal{X}$. Similarly, we have $H(X|Y) \geq R(X|Y)$. Note that Rényi entropy (like Shannon entropy) is always positive.

The following theorem is the main result of [2]:

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

$$H(G(X)|G) \geq R(G(X)|G) \geq r - \frac{2^{r-R(X)}}{\ln 2}.$$

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

This theorem clearly applies also to conditional probability distributions such as $P_{W|V=v}$ discussed above. If Eve's Rényi entropy $R(W|V=v)$ is known to be at least t and Alice and Bob choose $S = G(W)$ as their secret key, then

$$R(S|G, V=v) = R(G(W)|G, V=v) \geq r - 2^{r-t}/\ln 2.$$

The key S is indeed virtually secret because $H(S|G, V=v) \geq R(S|G, V=v)$ and hence $H(S|G, V=v)$ is arbitrarily close to maximal. More precisely, if $r < t$, then Eve's total information about S decreases exponentially in the excess compression $t-r$.

It should be pointed out that Theorem 1 cannot be generalized to Rényi entropy conditioned on a random variable, i.e., $R(G(W)|GV) \geq r - 2^{r-R(W|V)}/\ln 2$ is false in general [2].

3 The Effect of Side Information on Rényi Entropy

As described above, the reconciliation step consists of Alice and Bob exchanging suitable error-correction information U over the public channel. This information decreases Eve's Shannon entropy and usually also her Rényi entropy about W. For non-interactive reconciliation, Alice chooses an appropriate error-correction function f and sends $U = f(W)$ to Bob who can then reconstruct W with high probability from U and his prior knowledge YC.

The results of this paper will be derived for an arbitrary random variable X with probability distribution P_X and a side-information random variable U jointly distributed with X according to P_{XU}. However, they can just as well be applied to conditional distributions; our intended application is the key agreement scenario mentioned in the introduction, i.e., when P_X and $P_{X|U}$ are replaced by $P_{W|V=v}$ and $P_{W|V=v,U}$, respectively.

In general, giving side-information implies a reduction of entropy. Our goal is to derive upper bounds on the size of this reduction. Giving as side-information the fact that U takes on a particular value u, it is possible for both, Shannon and Rényi entropies, that the entropy increases or decreases. Moreover, the size of a reduction can be arbitrarily large. However, the expected reduction (for all values of U) of the Shannon entropy of X by giving U, called the mutual information between X and U, is bounded by $H(U)$:

$$H(X) - H(X|U) = I(X;U) \leq H(U) \tag{1}$$

which follows from the symmetry of $I(X;U)$ and the fact that Shannon entropy (conditional or not) is always positive.

Example 1 below illustrates two facts. First, the reduction of Rényi entropy implied by giving side-information $U = u$ can exceed the reduction of Shannon entropy, i.e.,

$$R(X) - R(X|U = u) > H(X) - H(X|U = u)$$

is possible. Second, it shows that the natural generalization of (1) to Rényi entropy, namely $R(X) - R(X|U) \leq R(U)$, is not true in general. However, Theorem 2 demonstrates that the weaker inequality $R(X) - R(X|U) \leq H(U)$ is always satisfied.

> *Example.* Let X be a random variable with alphabet $\mathcal{X} = \{a_1, \ldots, a_{10}, b_1, \ldots, b_{10}\}$, distributed according to $P_X(a_i) = 0.01$ and $P_X(b_i) = 0.09$ for $i = 1, \ldots, 10$. We have $H(X) \approx 3.79$ and $R(X) \approx 3.61$. Let $f : \mathcal{X} \to \{0, 1\}$ be defined as
>
> $$f(x) = \begin{cases} 0 & \text{if } x \in \{a_1, \ldots, a_9, b_{10}\} \\ 1 & \text{if } x \in \{a_{10}, b_1, \ldots, b_9\} \end{cases}$$
>
> and let $U = f(X)$. Then $H(X|U = 0) \approx 2.58$ and $R(X|U = 0) \approx 1.85$. The reduction of Rényi entropy when given $U = 0$ exceeds the reduction of Shannon entropy, i.e., $R(X) - R(X|U = 0) \approx 1.76$ whereas $H(X) - H(X|U = 0) \approx 1.21$.
>
> Because because f is deterministic, $H(U) = H(X) - H(X|U) \approx 0.69$. The expected entropy reductions are $H(X) - H(X|U) \approx 0.69$ and $R(X) - R(X|U) \approx 0.65$. Note that $R(U) \approx 0.50$ and that $R(X) - R(X|U)$ is indeed greater than $R(U)$ but less than $H(U)$.

$H(U)$ is not only the maximal expected decrease of Shannon entropy, but $H(U)$ is also an upper bound on the expected decrease of Rényi entropy, as the following theorem demonstrates.

Theorem 2. *Let X and U be two random variables with alphabets $\mathcal{X}$ and $\mathcal{U}$, respectively. The expected reduction of the Rényi entropy of X, when given U, does not exceed the Shannon entropy of U, i.e.,*

$$R(X) - R(X|U) \leq H(U),$$

with equality if and only if U is defined uniquely for each $x \in \mathcal{X}$ and P_U is the uniform distribution over $\mathcal{U}$ or a subset of $\mathcal{U}$.

Proof. The collision probability of X can be written as

$$\begin{aligned} P_c(X) &= \sum_{x\in\mathcal{X}} P_X(x)^2 \\ &= \sum_{x\in\mathcal{X}} \left(\sum_{u\in\mathcal{U}} P_{XU}(x, u) \right)^2 \\ &\geq \sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}} P_{XU}(x, u)^2 \\ &= \sum_{u\in\mathcal{U}} P_U(u)^2 \sum_{x\in\mathcal{X}} P_{X|U}(x, u)^2 \\ &= \sum_{u\in\mathcal{U}} P_U(u)^2 P_c(X|U = u), \end{aligned} \tag{2}$$

where the inequality follows from $(\sum_{i=1}^n p_i)^2 \geq \sum_{i=1}^n p_i^2$ for nonnegative p_i $(1 \leq i \leq n)$ and equality holds if and only if $p_i = 0$ for all but one i. Inserting (2) into the definition of Rényi entropy gives

$$\begin{aligned}
R(X) &= -\log P_c(X) \\
&\leq -\log\left(\sum_{u\in\mathcal{U}} P_U(u)^2 P_c(X|U=u)\right) \\
&= -\log\left(\sum_{u\in\mathcal{U}} P_U(u)\left[P_U(u)P_c(X|U=u)\right]\right) \\
&\leq -\sum_{u\in\mathcal{U}} P_U(u)\log\left[P_U(u)P_c(X|U=u)\right] \\
&= -\sum_{u\in\mathcal{U}} P_U(u)\left[\log P_U(u) + \log P_c(X|U=u)\right] \\
&= -\sum_{u\in\mathcal{U}} P_U(u)\log P_U(u) - \sum_{u\in\mathcal{U}} P_U(u)\log P_c(X|U=u) \\
&= H(U) + \sum_{u\in\mathcal{U}} P_U(u)\ R(X|U=u) \\
&= H(U) + R(X|U),
\end{aligned}$$

where the second inequality follows from Jensen's inequality [8] which holds with equality if and only if P_U is the uniform distribution over $\mathcal{U}$ or a subset of $\mathcal{U}$. □

In contrast to Shannon entropy, the expected Rényi entropy can *increase* when side information is revealed, i.e., $R(X) < R(X|U)$ is possible. This property is used in [2, 11] to prove that a key larger than suggested by Theorem 1 can be obtained by privacy amplification. The proof makes use of a conceptual oracle that is assumed to provide special side information U to Eve, called spoiling knowledge, which increases her Rényi entropy. Of course, this extra information cannot harm her because she could always discard it.

For example, if Eve has received $V = v$ over a binary symmetric channel, her Rényi entropy of the distribution $P_{W|V=v}$ is considerably lower than her Shannon entropy. But the oracle can increase her Rényi entropy with high probability almost to the Shannon entropy by telling her the number d of bits in v that she received incorrectly. Then, all strings at distance d of v are equally likely and Eve's Rényi entropy increases to the (new) Shannon entropy.

For a positive-valued random variable X, $E[X] \leq t$ implies that $P[X \geq kt] \leq 1/k$. Hence, according to Theorem 2, the probability that the leaking information $U = u$ decreases Rényi entropy by more than $kH(U)$ is at most $1/k$, i.e., $P[R(X) - R(X|U=u) \geq kH(U)] \leq 1/k$. However, such a high probability of partially exposing the string W is unacceptable in a key agreement scenario. The following theorem provides a much stronger result by showing that the above probability decreases in fact at most by $2\log|\mathcal{U}|$ except with negligible probability.

Theorem 3. *Let X and U be random variables with alphabets $\mathcal{X}$ and $\mathcal{U}$, respectively, and let $s > 0$ be an arbitrary security parameter. With probability at least $1 - 2^{-s}$, U takes on a value u for which*

$$R(X) - R(X|U=u) \ \leq\ 2\log|\mathcal{U}| + 2s.$$

Remark. The statement of the theorem is equivalent to

$$\sum_{u:\ R(X)-R(X|U=u)\leq 2\log|\mathcal{U}|+2s} P_U(u) \geq 1 - 2^{-s}.$$

Proof. We can bound the reduction only for those values u for which $P_U(u)$ is not too small. Let $\mathcal{E}$ be the event that $U \in \{u | P_U(u) \geq p_{min}\}$ for some given $p_{min} > 0$ and $p_{min} < \frac{1}{|\mathcal{U}|}$. There are $|\mathcal{U}|$ possible values u, and for at most $|\mathcal{U}| - 1$ values u, $P_U(u) < p_{min}$ can hold. Therefore we have $P[\overline{\mathcal{E}}] \leq |\mathcal{U}| \cdot p_{min}$ and $P[\mathcal{E}] \geq 1 - |\mathcal{U}| \cdot p_{min}$. If the event $\mathcal{E}$ occurs we have

$$\begin{aligned}
P_c(X|U=u) &= \sum_{x\in\mathcal{X}} P_{X|U}(x,u)^2 \\
&= \sum_{x\in\mathcal{X}} \left(\frac{P_{XU}(x,u)}{P_U(u)}\right)^2 \\
&\leq \sum_{x\in\mathcal{X}} \frac{P_X(x)^2}{P_U(u)^2} \\
&\leq \sum_{x\in\mathcal{X}} \frac{P_X(x)^2}{p_{min}^2} \\
&= \frac{P_c(X)}{p_{min}^2},
\end{aligned}$$

where we have made use of the fact that $P_{XU}(x,u) = P_X(x)P_{U|X}(u,x) \leq P_X(x)$. If we set $p_{min} = 2^{-s}/|\mathcal{U}|$, then $P_U(u) > p_{min}$ with probability $P[\mathcal{E}] \geq 1 - 2^{-s}$. The theorem now follows by taking logarithms on both sides of

$$P_c(X|U=u) \leq |\mathcal{U}|^2\, 2^{2s}\, P_c(X).$$

□

Because of its importance we restate Theorem 3 for the key-generation scenario, replacing P_X by $P_{W|V=v}$, with the side-information consisting of k bits, for instance k parity checks of W when W is an n-bit string.

Corollary 4. *Let W be a random variable with alphabet $\mathcal{W}$, let v and u be particular values of the correlated random variables V and U with alphabets $\mathcal{V}$ and $\mathcal{U}$, respectively, with $k = \log|\mathcal{U}|$, and let $s > 0$ be a given security parameter. Then, with probability at least $1 - 2^{-s}$, U takes on a value u such that the decrease in Rényi entropy by giving u,*

$$R(W|V=v) - R(W|V=v, U=u),$$

is at most $2k + 2s$.

4 Almost Uniform Distributions

As shown above, giving side information of the form $U = u$ can reduce the Rényi entropy by an arbitrary amount, although the probability that this happens is bounded by Theorem 3. In this section and the next we derive better bounds on the reduction for non-interactive reconciliation and special probability distributions. For uniform distributions and deterministic side-information $U = f(W)$ the reduction of Rényi entropy depends only on the size of the preimage of $u = f(x)$:

Lemma 5. *Let X be a random variable with alphabet $\mathcal{X}$, let $f : \mathcal{X} \to \mathcal{U}$ be an arbitrary function taking on values in a given set $\mathcal{U}$, let U be defined as $U = f(X)$, and set $\mathcal{X}_u = \{x \in \mathcal{X} : f(x) = u\}$. If X is distributed uniformly over $\mathcal{X}$, then*

$$R(X) - R(X|U=u) = \log\frac{|\mathcal{X}|}{|\mathcal{X}_u|}.$$

In particular, if f is such that $|\mathcal{X}_u|$ is the same for all $u \in \mathcal{U}$, knowledge of $U = u$ reduces the Rényi entropy by $\log |\mathcal{U}|$*.*

Proof. Because Rényi entropy equals Shannon entropy for the uniform distribution, we have $R(X) = \log |\mathcal{X}|$ and $R(X|U = u) = \log |\mathcal{X}_u|$, from which the first claim immediately follows. To prove the second claim, note that in this case $\frac{|\mathcal{X}|}{|\mathcal{X}_u|} = |\mathcal{U}|$ for all $u \in \mathcal{U}$. □

Theorems 6 and 7 state bounds on the reduction of Rényi entropy for almost uniform distributions. These results are applied in the next section to the analysis of the important class of scenarios where a given random experiment is repeated many times independently.

Theorem 6. *For given $\alpha > 1$ and $\beta > 1$, let X be a random variable with alphabet $\mathcal{X}$ and probability distribution P_X such that $\frac{1}{\alpha|\mathcal{X}|} \leq P_X(x) \leq \frac{\beta}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$. Define f, U and $\mathcal{X}_u$ as in Lemma 5. Then*

$$R(X) - R(X|U = u) \;\leq\; \log \frac{|\mathcal{X}|}{|\mathcal{X}_u|} + 4 \log \alpha + 2 \log \beta.$$

In particular, if f is such that $|\mathcal{X}_u|$ is the same for all $u \in \mathcal{U}$, then $R(X) - R(X|U = u) \leq \log |\mathcal{U}| + 4 \log \alpha + 2 \log \beta$.

Proof. We can bound $P_c(X)$ as follows:

$$P_c(X) \;=\; \sum_{x \in \mathcal{X}} P_X(x)^2 \;\geq\; |\mathcal{X}| \,\frac{1}{(\alpha\,|\mathcal{X}|)^2} \;=\; \frac{1}{\alpha^2\,|\mathcal{X}|}. \tag{3}$$

Using $P_U(u) \geq |\mathcal{X}_u| \frac{1}{\alpha|\mathcal{X}|}$ we get a similar upper bound for $P_c(X|U = u)$:

$$\begin{aligned} P_c(X|U = u) \;&=\; \sum_{x \in \mathcal{X}_u} P_{X|U}(x, u)^2 \\ &=\; \frac{1}{P_U(u)^2} \sum_{x \in \mathcal{X}_u} P_X(x)^2 \\ &\leq\; \left(\frac{\alpha\,|\mathcal{X}|}{|\mathcal{X}_u|}\right)^2 |\mathcal{X}_u| \left(\frac{\beta}{|\mathcal{X}|}\right)^2 \;=\; \frac{\alpha^2\beta^2}{|\mathcal{X}_u|}. \end{aligned} \tag{4}$$

Combining (3) and (4) gives

$$\frac{P_c(X|U = u)}{P_c(X)} \;\leq\; \frac{|\mathcal{X}|}{|\mathcal{X}_u|}\alpha^4\beta^2,$$

and the theorem follows by taking logarithms on both sides. □

The following theorem provides a tighter bound for distributions that are very close to uniform. In particluar, Theorem 7 is strictly tighter than Theorem 6 for $\gamma \leq 0.4563$. For $0 \leq \gamma \leq 0.3$ it is about 30% tighter.

Theorem 7. *For given $\gamma < \frac{1}{2}$, let X be a random variable with alphabet $\mathcal{X}$ and probability distribution P_X such that $\frac{1-\gamma}{|\mathcal{X}|} \leq P_X(x) \leq \frac{1+\gamma}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$. Define f, U and $\mathcal{X}_u$ as in Lemma 5. Then*

$$R(X) - R(X|U = u) \;\leq\; \log \frac{|\mathcal{X}|}{|\mathcal{X}_u|} + \log \frac{(1+\gamma)^2}{1-2\gamma}.$$

Proof. For each x, define δ_x as the deviation of $P_X(x)$ from the uniform distribution: $P_X(x) = \frac{1}{|\mathcal{X}|} + \delta_x$. Hence we have $|\delta_x| \leq \gamma/|\mathcal{X}|$, and $P_c(X|U=u)$ can be expressed as follows:

$$
\begin{aligned}
P_c(X|U=u) &= \sum_{x\in\mathcal{X}_u} P_{X|U}(x,u)^2 \\
&= \sum_{x\in\mathcal{X}_u} \left(\tfrac{P_{XU}(x,u)}{P_U(u)}\right)^2 \\
&= \frac{\sum_{x\in\mathcal{X}_u} P_X(x)^2}{P_U(u)^2} \\
&= \frac{\sum_{x\in\mathcal{X}_u} P_X(x)^2}{\left(\sum_{x\in\mathcal{X}_u} P_X(x)\right)^2} \\
&= \frac{\sum_{x\in\mathcal{X}_u} (\frac{1}{|\mathcal{X}|}+\delta_x)^2}{\left(\sum_{x\in\mathcal{X}_u} \frac{1}{|\mathcal{X}|} + \sum_{x\in\mathcal{X}_u}\delta_x\right)^2} \\
&= \frac{\frac{|\mathcal{X}_u|}{|\mathcal{X}|^2} + \frac{2}{|\mathcal{X}|}\sum_{x\in\mathcal{X}_u}\delta_x + \sum_{x\in\mathcal{X}_u}\delta_x^2}{\frac{|\mathcal{X}_u|^2}{|\mathcal{X}|^2} + 2\frac{|\mathcal{X}_u|}{|\mathcal{X}|}\sum_{x\in\mathcal{X}_u}\delta_x + \left(\sum_{x\in\mathcal{X}_u}\delta_x\right)^2} \\
&\leq \frac{\frac{|\mathcal{X}_u|}{|\mathcal{X}|^2} + \frac{2}{|\mathcal{X}|}|\mathcal{X}_u|\frac{\gamma}{|\mathcal{X}|} + |\mathcal{X}_u|\frac{\gamma^2}{|\mathcal{X}|^2}}{\frac{|\mathcal{X}_u|^2}{|\mathcal{X}|^2} - 2\frac{|\mathcal{X}_u|}{|\mathcal{X}|}|\mathcal{X}_u|\frac{\gamma}{|\mathcal{X}|}} \\
&= \frac{1+2\gamma+\gamma^2}{|\mathcal{X}_u| - 2\gamma|\mathcal{X}_u|}.
\end{aligned}
$$

In the third step we have made use of the fact that U is a deterministic function of X and thus $P_{XU}(x,u) = P_X(x)$. Using $P_c(X) \geq 1/|\mathcal{X}|$, we get

$$
\frac{P_c(X|U=u)}{P_c(X)} \leq \frac{|\mathcal{X}|}{|\mathcal{X}_u|} \cdot \frac{(1+\gamma)^2}{1-2\gamma},
$$

from which the theorem follows.

□

5 Independent Repetition of a Random Experiment

In many practical scenarios, a certain random experiment is repeated independently a large number of times. For example, W could be the result of receiving independently generated bits over a memoryless channel, as in the satellite scenario mentioned earlier. A fundamental theorem of information theory states that in such a scenario all occurring sequences can be divided into a typical set and a non-typical set, where the probability that a randomly selected sequence of length n lies in the typical set approaches 1 for all sufficiently large n. Furthermore, all sequences in the typical set are almost equally probable. This allows us to bound the decrease of Rényi entropy by the results of the last section.

In the following we will make use of strongly typical sequences [4]. Consider a probability distribution P_X over some finite set $\mathcal{X}$ where we assume $P_X(x) > 0$ for all $x \in \mathcal{X}$. Let $x^n = [x_1, \ldots, x_n]$ be a sequence of n digits of $\mathcal{X}$ and define $N_a(x^n)$ to be the number of occurrences of the symbol $a \in \mathcal{X}$ in the sequence x^n. A sequence $x^n \in \mathcal{X}^n$ is called ϵ-strongly typical if and only if

$$
(1-\epsilon)P_X(a) \leq \frac{N_a(x^n)}{n} \leq (1+\epsilon)P_X(a)
$$

for all $a \in \mathcal{X}$. Let $\mathcal{S}^n(\epsilon)$ be the set of all ϵ-strongly typical sequences of length n and define X^n to be a sequence of n independent and identically distributed random variables X_i with $P_{X_i} = P_X$ for $1 \leq i \leq n$. In other words, we have $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$. Let $o(n)$ be any function of n such that $\lim_{n\to\infty} o(n) = 0$. The following lemma asserts that, for sufficiently large n, the probability that $X^n \in \mathcal{S}^n(\epsilon)$ approaches 1 and that the cardinality of $\mathcal{S}^n(\epsilon)$ is close to $2^{nH(X)}$.

Lemma 8 ([4]). *Let X^n be a sequence of i.i.d. random variables distributed according to P_X. Then*

1. *For every $\delta > 0$, $P[X^n \in \mathcal{S}^n(\epsilon)] \geq 1 - \delta/n$, for sufficiently large n.*
2. *For all $x^n \in \mathcal{S}^n(\epsilon)$: $P_{X^n}(x^n) = 2^{-nH(X)+o(n)}$.*
3. *$|\mathcal{S}^n(\epsilon)| = 2^{nH(X)+o(n)}$.*

Because all sequences in $\mathcal{S}^n(\epsilon)$ are almost equally probable for sufficiently large n, the reduction of Rényi entropy is similar to the case of the uniform probability distribution where Rényi entropy behaves like Shannon entropy. This observation is stated as the next theorem.

Theorem 9. *Let X^n be a sequence of i.i.d. random variables distributed according to P_X, let $f : \mathcal{X}^n \to \mathcal{U}$ be an arbitrary function taking on values in a given set $\mathcal{U}$, and define $\mathcal{S}_u^n(\epsilon) = \{x^n \in \mathcal{S}^n(\epsilon) : f(x^n) = u\}$ and $U = f(X^n)$. For any $\delta > 0$ and sufficiently large n, the following holds with probability at least $1 - \delta/n$: X^n lies in $\mathcal{S}^n(\epsilon)$ and the reduction of Rényi entropy by giving $U = u$ is upper bounded by*

$$R(X^n) - R(X^n|U = u) \;\leq\; nH(X) - \log|\mathcal{S}_u^n(\epsilon)| + o(n).$$

In particular, if f is such that $|\{x \in \mathcal{X} : f(x) = u\}|$ is the same for all $u \in \mathcal{U}$ and $|\mathcal{U}| = 2^k$, then knowledge of $U = u$ reduces the Rényi entropy by at most $k + o(n)$.

Proof. By Lemma 8, $P_{X^n}(x^n) = 2^{-nH(X)+o(n)}$ for all $x^n \in \mathcal{S}^n(\epsilon)$, and $|\mathcal{S}^n(\epsilon)| = 2^{nH(X)+o(n)}$. Application of Theorem 6 with $\alpha = 2^{o(n)}$ and $\beta = 2^{o(n)}$ gives

$$\begin{aligned} R(X^n) - R(X^n|U = u) &\leq \log\frac{|\mathcal{S}^n(\epsilon)|}{|\mathcal{S}_u^n(\epsilon)|} + o(n) \\ &= nH(X) - \log|\mathcal{S}_u^n(\epsilon)| + o(n). \end{aligned}$$

□

Note that the second part of the theorem applies in particular to all linear functions such as parity checks from linear error correcting codes. Due to their wide-spread use linear error correcting codes are most likely to be used during the reconcilication phase. Theorem 9 can replace the spoiling knowledge argument in Maurer's proof [11] that the known results on secret key rate [10] hold also for a much stronger notion of secrecy.

6 Conclusions

The described link between information reconciliation and privacy amplification for unconditionally-secure secret-key agreement can be summarized as follows. Assume that Alice knows a random variable W and that Bob and Eve have partial knowledge about W, characterized by the random variables W' and V, respectively. These random variables could for instance result from the described satellite scenario with W and W' being functions of $[X, C]$ and $[Y, C]$, respectively, and with $V = [Z, C]$. In order to state the results in the strongest possible form we consider a particular value $V = v$ held by Eve rather than the average over all values of V.

When V gives less information than W' about W, i.e., $H(W|V) > H(W|W')$, and a lower bound $t > 0$ on the Rényi entropy of Eve's probability distribution of W is known, i.e., $R(W|V = v) \geq t$, then Alice and Bob can generate a shared secret key S as follows. Alice and Bob exchange error-correcting information U consisting of $k > H(W|W')$ bits over the public channel such that Bob can reconstruct W, i.e., $H(W|W'U) \approx 0$. Eve gains additional knowledge about W by seeing $U = u$. However, Corollary 4 shows that with probability at least $1 - 2^{-s}$ (over all values of U) where the security parameter s can be chosen arbitrarily, her Rényi entropy is bounded from below by $R(W|V = v, U = u) \geq t - 2k - 2s$. Using privacy amplification, Alice and Bob can now generate an r-bit secret key S, where r has to be chosen smaller than $t - 2k - 2s$ and Eve's total information about S is exponentially small in $t - 2k - 2s - r$, namely less than $2^{r-(t-2k-2s)}/\ln 2$ bits.

The main advantage of Theorem 3 is that it applies to any distribution and any reconciliation protocol whereas previously obtained results held only for particular distributions and protocols. However, as was demonstrated in Sections 4 and 5, a larger secret key than suggested by Theorem 3 can be obtained by Alice and Bob for special distributions. For instance, when Eve's information V consists of a long sequence of independent random variables with identical distribution, then the term $2k + 2s$ above can be replaced essentially by k if non-interactive reconciliation is used. It is conceivable that the bound of Theorem 3 can be tightened for other special distributions than those treated in Sections 4 and 5, and possibly even for general distributions.

Acknowledgement

It is a pleasure to thank Charles Bennett, Gilles Brassard, Claude Crépeau, and Martin Gander for interesting discussions and the anonymous referees for many helpful remarks.

References

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

[2] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, Nov. 1995. (To appear).

[3] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, pp. 210–229, Apr. 1988.

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

[5] G. Brassard, "A quantum jump in computer science," in *Computer Science Today* (J. van Leeuwen, ed.), vol. 1000 of *Lecture Notes in Computer Science*, Springer-Verlag, 1995.

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

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

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

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

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

[11] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides of One Tapestry* (R. E. Blahut *et al.*, eds.), Kluwer, 1994.

[12] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, vol. 1, (Berkeley), pp. 547–561, Univ. of Calif. Press, 1961.

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