A Counterexample to the Chain Rule for Conditional HILL Entropy*

And What Deniable Encryption Has to Do with It

Stephan Krenn[1,**], Krzysztof Pietrzak[2], and Akshay Wadia[3,***]

[1] IBM Research – Zurich, Rüschlikon
stephan.krenn@ist.ac.at
[2] Institute of Science and Technology Austria
pietrzak@ist.ac.at
[3] University of California, Los Angeles
awadia@cs.ucla.edu

Abstract. A chain rule for an entropy notion $H(\cdot)$ states that the entropy $H(X)$ of a variable X decreases by at most ℓ if conditioned on an ℓ-bit string A, i.e., $H(X|A) \geq H(X) - \ell$. More generally, it satisfies a chain rule for *conditional* entropy if $H(X|Y,A) \geq H(X|Y) - \ell$.

All natural information theoretic entropy notions we are aware of (like Shannon or min-entropy) satisfy some kind of chain rule for conditional entropy. Moreover, many *computational* entropy notions (like Yao entropy, unpredictability entropy and several variants of HILL entropy) satisfy the chain rule for conditional entropy, though here not only the *quantity* decreases by ℓ, but also the *quality* of the entropy decreases exponentially in ℓ. However, for the standard notion of conditional HILL entropy (the computational equivalent of min-entropy) the existence of such a rule was unknown so far.

In this paper, we prove that for conditional HILL entropy no meaningful chain rule exists, assuming the existence of one-way permutations: there exist distributions X, Y, A, where A is a distribution over a *single* bit, but $H^{\mathsf{HILL}}(X|Y) \gg H^{\mathsf{HILL}}(X|Y,A)$, even if we simultaneously allow for a massive degradation in the quality of the entropy.

The idea underlying our construction is based on a surprising connection between the chain rule for HILL entropy and deniable encryption.

Keywords: Computational entropy, HILL entropy, Conditional chain rule.

1 Introduction

Various information theoretic entropy notions are used to quantify the amount of randomness of a probability distribution. The most common one is Shannon

* This work was partly funded by the European Research Council under an ERC Starting Grant (259668-PSPC).

** This work was done while the author was at IST Austria.

*** This work was done while the author was visiting IST Austria.

A. Sahai (Ed.): TCC 2013, LNCS 7785, pp. 23–39, 2013.

entropy, which measures the incompressibility of a distribution. In cryptographic settings the notion of min-entropy, measuring the unpredictability of a random variable, is often more convenient to work with.

One of the most useful tools for manipulating and arguing about entropies are chain rules, which come in many different flavors for different entropy notions. Roughly, a chain rule captures the fact that the entropy of a variable X decreases by at most the entropy of another variable A if conditioned on A. For Shannon entropy, we have a particularly simple chain rule

$$H(X|A) = H(X, A) - H(A)$$

More generally, one can give chain rules for conditional entropies by considering the case where X has some entropy conditioned on Y, and bound by how much the entropy drops when given A. The chain rule for Shannon entropy naturally extends to this case

$$H(X|Y, A) = H(X|Y) - H(A)$$

For min-entropy (cf. Definition 2.1) an elegant chain rule holds if one uses the right notion of conditional min-entropy. The worst case definition $H_\infty(X|Y) = \min_y H_\infty(X|Y = y)$ is often too pessimistic. An average-case notion has been defined by [5] (cf. Definition 2.2), and they show it satisfies the following chain rules ($H_0(A)$ is the logarithm of the size of the support of A):

$$\widetilde{H}_\infty(X|A) \geq H_\infty(X) - H_0(A) \quad \text{and} \quad \widetilde{H}_\infty(X|Y, A) \geq \widetilde{H}_\infty(X|Y) - H_0(A)\,.$$

1.1 Computational Entropy

The classical information theoretic notions anticipate computationally unbounded parties, e.g. no algorithm can compress a distribution below its Shannon entropy and no algorithm can predict it better than exponentially in its min-entropy. Under computational assumptions, in particular in cryptographic settings, one can talk about distribution that appear to have high entropy only for computationally bounded parties. The most basic example are pseudorandom distributions, where $X \in \{0,1\}^n$ is said to be pseudorandom if it cannot be distinguished from the uniform distribution U_n by polynomial size distinguishers. So X appears to have n bits of Shannon and n bits of min-entropy.

Pseudorandomness is a very elegant and tremendously useful notion, but sometimes one has to deal with distributions which do not look uniform, but only seem to have some kind high entropy. Some of the most prominent such notions are HILL, Yao and unpredictability entropy. Informally, a distribution X has k bits of HILL-pseudoentropy [13] (conditioned on Z), if cannot be distinguished from some variable Y with k bits of min-entropy (given Z). X has k bits of Yao entropy [1,20] (conditioned on Z) if it cannot be compressed below k bits (given Z), and X has k bits of unpredictability entropy [14] conditioned on Z if no efficient adversary can guess X better than with probability 2^{-k} given Z.[1] When we talk about, say the HILL entropy of X, not only its *quantity* k is of

[1] Unlike HILL and Yao, unpredictability entropy is only interesting if the conditional part Z is not empty, otherwise it coincides with min-entropy.

interest, but also its *quality* which specifies against what kind of distinguishers X looks like having k bits of min-entropy. This is specified by giving two additional parameters (ε, s), and the meaning of $H^{\mathsf{HILL}}_{\varepsilon,s}(X) = k$ is that X cannot be distinguished from some Y with min-entropy k by distinguishers of size s with advantage greater than ε.

Chain rules for (conditional) entropy are easily seen to hold for some computational entropy notions (in particular for (conditional) Yao and unpredictability), albeit there are two caveats. First, one must typically assume that the part A we condition on comes from an efficiently samplable distribution, we will always set $A \in \{0,1\}^\ell$. Second, the quality of the entropy (the distinguishing advantage, circuit size, or both) typically degrades exponentially in ℓ. The chain rules for (conditional) computational entropy notions H we know state that for any distribution (X, Y, A) where $A \in \{0,1\}^\ell$ (X, Y, A) where $A \in \{0,1\}^\ell$

$$H_{\varepsilon',s'}(X|Y,A) \geq H_{\varepsilon,s}(X|Y) - \ell \tag{1}$$

where $\varepsilon' = \mu(\varepsilon, 2^\ell)$, $s' = s/\nu(2^\ell, \varepsilon)$ for some polynomial functions μ, ν. For HILL entropy such a chain rule has only recently been found [7,15] (cf. Lemma 2.6), but only holds for the unconditional case, i.e., when Y in (1) is empty (or at least very short, cf. Theorem 3.7 [9]). Whether or not a chain rule holds for conditional HILL has been open up to now. In this paper we give a counterexample showing that the chain rule for conditional HILL entropy does not hold in a very strong sense.

We will not try to formally define what constitutes a chain rule for a computational entropy notion, not even for the special case of HILL entropy we consider here, as this would seem arbitrary. Instead, we will specify what it means that conditional HILL entropy does not satisfy a chain rule. This requirement is so demanding that it leaves little room for any kind of meaningful positive statement that could be considered as a chain rule.

We will say that an ensemble of distributions $\{(X_n, Y_n, A_n)\}_{n\in\mathbb{N}}$ forms a counterexample to the chain rule for conditional HILL entropy if

- X_n has a lot of high quality HILL entropy conditioned on Y_n : that is, $H^{\mathsf{HILL}}_{\epsilon,s}(X_n|Y_n) = z_n$ where (high quantity) $z_n = n^\alpha$ for some $\alpha > 0$ (we will achieve any $\alpha < 1$) and (high quality) for every polynomial $s = s(n)$ we can set $\epsilon = \epsilon(n)$ to be negligible.
- The HILL entropy of X_n drops by a constant fraction conditioned additionally on a single bit $A_n \in \{0,1\}$, even if we only ask for very low quality entropy: (large quantitative gap) $H^{\mathsf{HILL}}_{\epsilon',s'}(X_n|Y_n, A_n) < \beta \cdot H^{\mathsf{HILL}}_{\epsilon,s}(X_n|Y_n)$ for $\beta < 1$ (we achieve $\beta < 0.6$) and (low quality) $\epsilon' > 0$ is constant (we achieve any $\epsilon' < 1$) and $s' = s'(n)$ is a fixed polynomial.

Assuming the existence of one-way permutations, we construct such an ensemble of distributions $\{(X_n, Y_n, A_n)\}_{n\in\mathbb{N}}$ over $\{0,1\}^{1.5n^2} \times \{0,1\}^{3n^2} \times \{0,1\}$.

$$H^{\mathsf{HILL}}_{\varepsilon',s'}(X_n|Y_n, A_n) \;<\; H^{\mathsf{HILL}}_{\varepsilon,s}(X_n|Y_n) - 1.25n$$

Moreover $H^{\mathsf{HILL}}_{\varepsilon,s}(X|Y) \approx 3n$, which gives a multiplicative gap of $(3n - 1.25n)/3n < 0.6$

$$H^{\mathsf{HILL}}_{\varepsilon',s'}(X_n|Y_n, A_n) \;<\; 0.6 \cdot H^{\mathsf{HILL}}_{\varepsilon,s}(X_n|Y_n)\,,$$

where $H^{\mathsf{HILL}}_{\varepsilon,s}$ is high-quality cryptographic-strength pseudoentropy (i.e., for any polynomial $s = s(n)$ we can choose $\varepsilon = \varepsilon(n)$ to be negligible) and (ε', s') is extremely low end where ε' can be any constant < 1 and s is a fixed polynomial (depending only the complexity of evaluating the one-way permutation). The entropy gap $1.25n$ we achieve is constant factor of entire HILL entropy $H^{\mathsf{HILL}}_{\varepsilon,s}(X_n|Y_n) \approx 3n$ in X. The gap is roughly the square root of the length $m = 4.5n^2$ of the variables (X_n, Y_n). This can be easily increased from $n \approx m^{1/2}$ to $n \approx m^{1-\gamma}$ for any $\gamma > 0$.

Interestingly, for several variants of conditional HILL entropy, chain rules in the conditional case do hold. In particular, this is the case for the so called *decomposable*, *relaxed* and *simulatable* versions of HILL entropy (cf. [9] and references therein).

1.2 Counterexamples from Deniable Encryption and One-Way Permutations

Deniable encryption has been proposed in 1997 by Canetti et al. [3], if such schemes actually exists has been an intriguing open problem ever since. The only known negative result is due to Bendlin et al. [2] who show that *receiver* deniable *non-interactive* public-key encryption is impossible. Informally, a *sender* deniable public-key encryption scheme (we will just consider bit-encryption) is a semantically secure public-key encryption scheme, which additionally provides some efficient way for the sender of a ciphertext C computed as $C := enc(pk, B, R)$ to come up with some *fake* randomness R' which explains C as a ciphertext for the opposite message $1 - B$. That is $C = enc(pk, 1 - B, R')$, and for a random B, (C, B, R) and $(C, 1 - B, R')$ are indistinguishable.

We show a close connection between deniable encryption and HILL entropy: any deniable encryption scheme provides a counterexample to the chain rule for conditional HILL entropy. This connection has been the starting point for the counterexample constructed in this paper. Unfortunately, this connection does not immediately prove the impossibility of a chain rule, as deniable encryption is not known to exist. Yet, a closer look shows that we do not need all the functionalities of deniable encryption to construct a counterexample. In particular, neither the faking algorithm nor decryption must be efficient. We will exploit this to get a counterexample from any one-way permutation.

1.3 Related Work

The concept of HILL entropy has first been introduced by Håstad et al. [13], and the conditional variant was suggested by Hsiao et al. [14]. Other notions of computational entropy include Yao entropy [1,20], unpredictability entropy [14], and metric entropy [1].

Chain rules for these entropy notions are known, e.g., Fuller et al. [8] for metric entropy, where they also show a connection between metric entropy and deterministic encryption. A chain rule for HILL entropy was proved independently by Reingold et al. [15] (it is a corollary of the more general dense model theorem proven in this work) and Dziembowski and Pietrzak [7] (as a tool for proving security of leakage-resilient cryptosystems). This chain rule only applies in the unconditional setting, but for some variants of HILL entropy, chain rules are known in the conditional setting as well. Chung et al. [4] proved a chain rule for *samplable* HILL entropy, a variant of HILL entropy where one requires the high min-entropy distribution Y as in Definition 2.5 to be efficiently samplable. Fuller et al. [8] give a chain rule for *decomposable* metric entropy (which implies HILL entropy). Reyzin [16] (cf. Theorem 2 and the paragraph following it in [16]) gives a chain rule for conditional *relaxed* HILL entropy, such a rule is implicit in the work of Gentry and Wichs [10].

A chain rule for normal conditional HILL entropy (citing [8]) "remains an interesting open problem". The intuition underlying the counterexample we construct (giving a negative answer to this open problem) borrows ideas from the deniable encryption scheme of Dürmuth and Freeman [6] presented at Eurocrypt 2011, which unfortunately later turned out to have a subtle flaw. In their protocol, after receiving the ciphertext, the receiver (knowing the secret key) helps the sender to evaluate a faking algorithm by sending some information the sender could not compute efficiently on its own. It is this interactive phase that is flawed. However, it turns out that for our counterexample to work, the faking algorithm does not need to be efficiently computable, and thus we can already use the first part of their protocol as a counterexample. Moreover, as we don't require an efficient decryption algorithm either, we can further weaken our assumptions and base our construction on any one-way permutation instead of trapdoor permutations.

1.4 Roadmap

This document in structured as follows: in Section 2 we recap the basic definitions required for paper. In Section 3 we then give the intuition underlying our results by deriving a counterexample to the chain rule for conditional HILL entropy from any sender-deniable bit-encryption scheme. The counterexample based on one-way permutations is then formally presented in Section 4.

2 Preliminaries

In this section we recap the basic definitions required for this document. We start by defining some standard notation, and then recapitulate the required background of entropy measures, hardcore predicates, and Stirling's formula.

We say that $f(n) = \mathcal{O}(g(n))$, if $f(n)$ is asymptotically bounded above by $g(n)$, i.e., there exists a $k \in \mathbb{N}$ such that $|f(n)| \leq k|g(n)|$ for all $n > k$. Similarly, $f(n) = \omega(g(n))$, if $f(n)$ asymptotically dominates $g(n)$, i.e., for every $k \in \mathbb{N}$,

there exists $n_k \in \mathbb{N}$, such that for all $n > n_k$ we have that $kg(n) < f(n)$. A function $\nu(n)$ is called *negligible*, if it vanishes faster than every polynomial, i.e., for every integer k, there exists an integer n_k such that $\nu(n) < n^{-k}$ for all $n > n_k$, or alternatively, if $n^{-k} = \omega(\nu(n))$ for all k.

By $|\mathcal{S}|$ we denote the cardinality of some set $\mathcal{S}$. We further write $s \xleftarrow{\$} \mathcal{S}$ to denote that s is drawn uniformly at random from $\mathcal{S}$. The *support* of a probability distribution X, denoted by $\mathrm{supp}(X)$, is the set of elements to which X assigns non-zero probability mass, i.e., $\mathrm{supp}(X) = \{x \mid \mathsf{Pr}[X = x] > 0\}$. A distribution X is called *flat*, if it is uniform on its support, i.e., $\forall x \in \mathrm{supp}(X), \mathsf{Pr}[X = x] = 1/|\mathrm{supp}(X)|$. Finally, we use the notation $\mathsf{Pr}[\mathcal{E} : \Omega]$ to denote the probability of event $\mathcal{E}$ over the probability space Ω. For example, $\mathsf{Pr}\left[f(x) = 1 : x \xleftarrow{\$} \{0,1\}^n\right]$ is the probability that $f(x) = 1$ for a uniformly drawn x in $\{0,1\}^n$.

2.1 Entropy Measures

Informally, the entropy of a random variable X is a measure of the uncertainty of X. In the following we define those notions of entropy required for the rest of the paper.

Min-Entropy. Min-entropy is often useful in cryptography, as it ensures that the success probability of even a computationally unbounded adversary guessing the value of a sample from X is bounded above by $2^{-H_\infty(X)}$:

Definition 2.1 (Min-Entropy). *A random variable X has* min-entropy *k, denoted by $H_\infty(X) = k$, if*

$$\max_x \mathsf{Pr}[X = x] = 2^{-k}.$$

While a conditional version of min-entropy is straightforward to formulate, Dodis et al. [5] introduced the notion of *average min-entropy*, which is useful, if the adversary does not have control over the variable one is conditioning on.

Definition 2.2 (Average min-Entropy). *For a pair (X, Z) of random variables, the* average min-entropy *of X conditioned on Z is*

$$\widetilde{H}_\infty(X|Z) = -\log \mathop{\mathbb{E}}_{z \leftarrow Z} \max_x \mathsf{Pr}[X = x|Z = z] = -\log \mathop{\mathbb{E}}_{z \leftarrow Z} 2^{-H_\infty(X|Z=z)},$$

where the expectation is over all z with non-zero probability.

Similarly to min-entropy, an adversary learning Z can only predict X with probability $2^{-\widetilde{H}_\infty(X|Z)}$.

HILL Entropy. While min-entropy guarantees an information-theoretic bound on the probability of an adversary guessing a random variable, this bound might not be reached by any adversary of a limited size. For instance, this is the case for pseudorandom distributions. This fact is taken into account in computational variants of entropy.

Before formally defining HILL entropy, the computational equivalent of min-entropy, we recap what it means for two probability distributions to be close in a computational sense:

Definition 2.3 (Closeness of Distributions). *Two probability distributions X and Y are (ε, s)-close, denoted by $X \sim_{\varepsilon,s} Y$, if for every circuit D of size at most s the following holds:*

$$|\mathsf{Pr}[\mathsf{D}(X) = 1] - \mathsf{Pr}[\mathsf{D}(Y) = 1]| \leq \varepsilon\,.$$

We further say that two ensembles of distributions $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are $\varepsilon(n)$-computationally-indistinguishable if for every positive polynomial $\mathsf{poly}(n)$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, it holds that $X_n \sim_{\varepsilon(n),\mathsf{poly}(n)} Y_n$.

Informally, a random variable X has a high HILL entropy, if it is computationally indistinguishable from a random variable with high min-entropy, cf. Håstad et al. [13]:

Definition 2.4 (HILL Entropy). *A distribution X has* HILL entropy *k, denoted by $H^{\mathsf{HILL}}_{\varepsilon,s}(X) \geq k$, if there exists a distribution Y satisfying $H_\infty(Y) \geq k$ and $X \sim_{\varepsilon,s} Y$.*

Intuitively, in the above definition, k can be thought of as the quantity of entropy in X, whereas ε and s specify its quality: the larger s and the smaller ε, the closer X is to a random variable Y with information-theoretic min-entropy k in a computational sense.

A conditional version of HILL entropy can be defined similarly as a computational analogue to average min-entropy [14]:

Definition 2.5 (Conditional HILL Entropy). *Let X, Z be random variables. X has* conditional HILL entropy *$H^{\mathsf{HILL}}_{\varepsilon,s}(X|Z) \geq k$ conditioned on Z, if there exists a collection of distributions $\{Y_z\}_{z\in Z}$ giving rise to a joint distribution (Y, Z) such that $\widetilde{H}_\infty(Y|Z) \geq k$, and $(X, Z) \sim_{\varepsilon,s} (Y, Z)$.*

It has been shown that conditioning X on a random variable of length at most ℓ reduces the HILL entropy by at most ℓ bits, if the quality may decrease exponentially in ℓ [7,15,8]:

Lemma 2.6 (Chain Rule for HILL Entropy). *For a random variable X and $A \in \{0,1\}^\ell$ it holds that*

$$H^{\mathsf{HILL}}_{\varepsilon',s'}(X|A) \geq H^{\mathsf{HILL}}_{\varepsilon,s}(X) - \ell\,,$$

where $\varepsilon' \approx 2^\ell\varepsilon$ and $s' \approx s\varepsilon'^2$.

2.2 Hardcore Predicates

The counterexample we present in Section 4 is based on the existence of one-way permutations, which we define next. Intuitively, a permutation is one-way, if it is easy to compute but hard to invert. For an extensive discussion, see [11, Chapter 2]. The following definition is from [19]:

Definition 2.7 (One-Way Permutation). *A length-preserving function* $\pi : \{0,1\}^* \to \{0,1\}^*$ *is called a* one-way permutation*, if* π *is computable in polynomial time, if for every* n*,* π *restricted to* $\{0,1\}^n$ *is a permutation, and if for every probabilistic polynomial-time algorithm* A *there is a negligible function* ν *such that the following holds:*

$$\mathsf{Pr}\left[\mathsf{A}(\pi(x)) = x \,:\, x \xleftarrow{\$} \{0,1\}^n\right] < \nu(n)\,.$$

While for a one-way permutation, given $\pi(x)$ it is hard to compute x in its entirety, it may be easy to efficiently compute a large fraction of x. However, for our construction we will need that some parts of x cannot be computed with better probability than by guessing. This is captured by the notion of a hardcore predicate [12]. We use the formalization from [18]:

Definition 2.8 (Hardcore Predicate). *We call* $p : \{0,1\}^* \to \{0,1\}$ *a* $(\sigma(n), \nu(n))$-hardcore predicate *for a one-way permutation* π*, if it is efficiently computable, and if for every adversary running in at most* $\sigma(n)$ *steps, the following holds:*

$$\mathsf{Pr}\left[\mathsf{A}(\pi(x)) = p(x) \,:\, x \xleftarrow{\$} \{0,1\}^n\right] < \frac{1}{2} + \nu(n)\,.$$

It is well known that a one-way permutation π with a hardcore predicate p can be derived from any one-way permutation π' as follows [12]: for r of the same length as x, define $\pi(x,r) := (\pi'(x), r)$ and $p(x,r) := \langle x, r\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product modulo 2.

2.3 Stirling's Formula

Stirling's approximation [17] states that for any integer n it holds that:

$$\log n! = n\log n - \frac{n}{\ln 2} + \mathcal{O}(\log n)\,.$$

In our results we will make use of the following lemma, which directly follows from Stirling's formula.

Lemma 2.9. *For every integer* $a > 1$ *we have that*

$$\log\binom{an}{n} = an\log a - (a-1)n\log(a-1) + \mathcal{O}(\log n)\,. \tag{2}$$

3 A Counterexample from Sender Deniable Encryption

We start this section by defining sender deniable encryption schemes, and then show how such a scheme leads to a counterexample to the chain rule for conditional HILL entropy.

As the existence of sender deniable public key encryption schemes is an open problem, this implication does not directly falsify the chain rule. However, it shows up an interesting connection, and gives the idea underlying our result, as the proof given in Section 4 was strongly inspired by deniable encryption. We stress that the main purpose of this section is to give the reader some intuition, and thus we do not fully formalize all steps here.

3.1 Sender Deniable PKE

Deniable encryption, first introduced by Canetti et al. [3], is a cryptographic primitive offering protection against *coercion*. Consider therefore the following scenario: a sender sends an encrypted message to a receiver over a public channel. After the transmission, an adversary who wishes to learn the message sent, coerces one of the parties into revealing the secret information that was used to run the protocol (i.e., the secret message, the random tape used to generate keys, etc.). If the parties used a semantically secure but non-deniable encryption scheme, the adversary can check the consistency of the protocol transcript (which was carried over a public channel) and the secret information of the party, in particular learning whether the provided message was indeed the one being encrypted. A deniable encryption scheme tackles this problem by providing a *faking* algorithm. The faking algorithm allows a coerced party to come up with fake keys and random tapes that, while being consistent with the public transcript, correspond to an arbitrary message different from the real one. Deniable encryption schemes are classified as *sender deniable*, *receiver deniable* or *bi-deniable*, depending on which party can withstand coercion. For our purposes, we will focus only on sender deniable encryption schemes.

We will think of an encryption scheme as a two-party protocol between a sender S and a receiver R. The sender's input as well as the receiver's output are messages m from a message space M. For an encryption protocol ψ, we will denote by $tr_\psi(m, r_S, r_R)$ the (public) transcript of the protocol, where m is the sender's input, and r_S and r_R are the sender's and the receiver's random tapes, respectively. Let $tr_\psi(m)$ be the random variable distributed as $tr_\psi(m, r_S, r_R)$ where r_S and r_R are uniformly picked in their supports. A sender deniable encryption scheme is then defines as follows [3]:

Definition 3.1 (Sender Deniable PKE). *A protocol ψ with sender S and receiver R, and security parameter n, is a $\delta(n)$-sender-deniable encryption protocol if:*

Correctness: *The probability that R's output is different from S's input is negligible (as a function of n).*

Security: *For every $m_1, m_2 \in M$, the distributions $tr_\psi(m_1)$ and $tr_\psi(m_2)$ are computationally indistinguishable.*

Deniability: *There exists an efficient* faking algorithm ϕ *having the following property with respect to any $m_1, m_2 \in M$. Let r_S, r_R be uniformly chosen random tapes for S and R, respectively, let $c = tr_\psi(m_1, r_S, r_R)$, and let $\bar{r_S} = \phi(m_1, r_S, c, m_2)$. Then the random variables*

$$(m_2, \bar{r_S}, c) \text{ and } (m_2, r'_S, tr_\psi(m_2, r'_S, r'_R))$$

are $\delta(n)$-computationally-indistinguishable, where r'_S and r'_R are independent, uniformly chosen random tapes for S and R.

For notational convenience, when considering bit-encryption schemes (i.e., $M = \{0,1\}$), we will ignore the last argument of the algorithm ϕ. Further, we will call a scheme negl-sender-deniable if $\delta(n)$ is some negligible function in n.

Canetti et al. [3] give a construction of sender deniable encryption with $\delta(n) = 1/\mathsf{poly}(n)$ for some polynomial $\mathsf{poly}(n)$. However, the problem of constructing a sender deniable scheme with a negligible $\delta(n)$ has remained open since (recently, Dürmuth and Freeman [6] proposed a construction of negl-sender-deniable encryption scheme, but their proof was found to be flawed, cf. the foreword of the full version of their paper).

3.2 A Counterexample from Deniable Encryption

In the following we explain how a non-interactive negl-sender-deniable encryption scheme for message space $M = \{0,1\}$ would lead to a counterexample to the chain rule for conditional HILL entropy. Let ψ be the encryption algorithm of this scheme.

Let B be a uniformly random bit, and let R_S be the uniform distribution of appropriate length that serves as the random tape of the sender. Over this space, we now define the following random variables:

- Let C be a ciphertext, i.e., $C := \psi(B, R_S)$.
- Let R'_S be the *fake* random tapes for the sender, i.e.,

$$R'_S := \phi(B, R_S, C)$$

Fix now a transcript c, and let b_c be the bit that the receiver outputs for c. We then define the sets R_c and R'_c as follows:

$$R_c := \{r_S \mid c = \psi(b_c, r_S)\},$$
$$R'_c := \{\phi(b_c, r_S, c) \mid r_S \in R_c\}.$$

Note that for every $r'_S \in R'_c$, we have that $c = \psi(1 - b_c, r'_S)$.

In the following we will make two simplifying assumptions about the encryption scheme. We note that we make these assumptions only for the sake of presentation. The subsequent arguments can still be adapted to work without them:

(i) Firstly, for all public keys and all ciphertexts c_1, c_2, we have that $|R_{c_1}| = |R_{c_2}|$ and $|R'_{c_1}| = |R'_{c_2}|$. We will call these cardinalities $|R|$ and $|R'|$, respectively. Put differently, we assume that $|R|$ and $|R'|$ only depend on the security parameter n.

(ii) Secondly, we assume that ϕ induces a flat distribution on R'_c, i.e., if Z is the conditional distribution on R_c given c, then $\phi(b_c, Z)$ is flat on R'_c.

We now argue that the gap between $H^{\mathsf{HILL}}_{\varepsilon,s}(R'_S|C)$ and $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'_S|C,B)$ is very large.[2]

1. The deniability property implies that no PPT adversary can distinguish between real and fake random tapes for the sender. Thus, the distributions (R_S, C) ad (R'_S, C) are computationally indistinguishable. Therefore,
$$H^{\mathsf{HILL}}_{\varepsilon,s}(R'_S|C) \geq \widetilde{H}_\infty(R_S|C) = \log(|R|).$$
2. Now consider $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'_S|C,B)$. We argue that this value is bounded above by (roughly) $\widetilde{H}_\infty(R'_S|C,B)$. This is because given ciphertext c *and* bit b, there exists an efficient test to check if $r \in \mathrm{supp}(R'_S)$ or not. Indeed, given a random tape r, a transcript c and bit b, we can check if r is in the support of R'_S or not as follows: run the sender in ψ with input $1-b$ and random tape r. The resulting ciphertext is equal to c, if and only if r lies in the support of R'_S. Thus, for any distribution Z such that (R'_S, C, B) and (Z, C, B) are computationally indistinguishable, it must be the case that the support of Z is (almost) a subset of the support of R'_S. Using further that R'_S is flat, we get that:
$$H^{\mathsf{HILL}}_{\varepsilon',s'}(R'_S|C,B) \approx \widetilde{H}_\infty(R'_S|C) = \log(|R'|)\,.$$
3. To complete the argument, we need to show that the difference between $\log(|R|)$ and $\log(|R'|)$ is large. We do so by relating this difference to the decryption error of the encryption scheme. Consider a ciphertext c that decrypts to bit b. Consider the set of all random tapes that produce this ciphertext c. Out of these, $|R_c|$ of them encrypt bit b to c, while $|R'_c|$ of them encrypt bit $1-b$ to c. Thus, an error will be made in decrypting c when the sender wanted to encrypt bit $1-b$, but picked its random tape from the set R'_c. Combining this observation with the simplifying assumptions made earlier, we get that the decryption error of the encryption scheme is given by $\frac{|R'|}{|R|+|R'|}$. As the decryption error is negligible by Definition 3.1, we obtain that:
$$\log(|R|) - \log(|R'|) = \omega(\log(n))\,.$$

Combining the above arguments yields that the difference between $H^{\mathsf{HILL}}_{\varepsilon,s}(R_S|C)$ and $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'_S|C,B)$ is at least super-logarithmic in the security parameter of the encryption scheme.

[2] For clarity of exposition, we will not detail the relation of the parameters ε, s and ε', s' in this section. The counterexample in Section 4 gives a formal treatment of all parameters, though. Furthermore, we do not make the public key explicit in the conditional entropies in the following.

4 Disproving the Conditional Chain Rule

In the previous section we showed that the existence of sender-deniable bit encryption schemes would disprove the chain rule for conditional HILL entropy. However, the existence of such schemes is currently unknown. Thus, in this section we give a counterexample which only relies on the existence of one-way permutations.

In the following we let $\pi : \{0,1\}^* \to \{0,1\}^*$ be a one-way permutation with hardcore predicate $p : \{0,1\}^* \to \{0,1\}$. Furthermore, we define the probabilistic algorithm C, taking a bit b and a parameter n in unary as inputs, as follows:

- C draws $3n$ distinct elements $x_1, \ldots, x_{3n} \xleftarrow{\$} \{0,1\}^n$ such that $p(x_i) = b$ for $1 \leq i \leq 2n$ and $p(x_j) = 1 - b$ for $2n < j \leq 3n$.
- C outputs $\pi(x_1), \ldots, \pi(x_{3n})$ in lexicographical order.

We now define two random variables R and R' conditioned on a value $c = \mathsf{C}(1^n, b)$ as $1.5n$-tuple in $\{0,1\}^n$ as follows:

R consists of	R' consists of
– a uniformly random subset of $x_1, \ldots, x_{2n}$ of cardinality n, and – a uniformly random subset of $x_{2n+1}, \ldots, x_{3n}$ of cardinality $n/2$, in lexicographical order.	– a uniformly random subset of $x_1, \ldots, x_{2n}$ of cardinality $n/2$, and – $x_{2n+1}, \ldots, x_{3n}$, in lexicographical order.

Having said this, we can now state the main result of this paper. Informally, it says that R' conditioned on C has high HILL entropy of high quality, while additionally conditioning on the single bit B decreases both, quantity *and* quality of the entropy by factors polynomial in n:

Theorem 4.1 (Counterexample for a Conditional Chain Rule). *Let p be a $(\sigma(n), \nu(n))$-hardcore predicate for π, and let $B \xleftarrow{\$} \{0,1\}$ and $C = \mathsf{C}(1^n, B)$. Then for all sufficiently large n it holds that:*

$$H^{\mathsf{HILL}}_{\varepsilon,s}(R'|C) - H^{\mathsf{HILL}}_{\varepsilon',s'}(R'|C,B) > \frac{5}{4}n\,,$$

where

$$\varepsilon(n) = n\nu(n), \qquad \varepsilon'(n) = 0.99,$$
$$s(n) = \sigma(n) - \mathcal{O}(n(\sigma_p(n) + \sigma_\pi(n)), \qquad s'(n) = 1.5n(\sigma_p(n) + \sigma_\pi(n)),$$

where $\sigma_p(n)$ and $\sigma_\pi(n)$ denote the required running times to evaluate p and π, respectively, on n-bit inputs.

We now briefly want to discuss what the theorem means for the potential loss of quality and quantity of conditional HILL entropy.

Loss in **Quality** *of Entropy.* Note that ε and s are roughly of the same size as the security parameters of p, while ε' and s' are completely independent thereof. This means that even if we have $(\sigma(n), \nu(n)) = (\mathsf{poly}_1(n), 1/\mathsf{poly}_2(n))$ for some polynomials $\mathsf{poly}_i(n)$, $i = 1, 2$, as is the case for cryptographic hardcore predicates, the loss of neither of the parameters can be bounded above by a constant, but is polynomial in n.

Loss in **Quantity** *of Entropy.* Despite this large tolerated loss in the quality of the entropy, Theorem 4.1 says that conditioning on a single bit of extra information can still decrease the conditional HILL entropy by arbitrarily large additive factors by choosing n sufficiently large.

Together this implies that in order to formulate a chain rule for conditional HILL entropy, neither the loss in quality nor in quantity could be bounded by a constant, as would be desirable for a reasonable such rule, but must also depend on the size of the random variable R' whose entropy one wants to compute.

4.1 Proof of Theorem 4.1

Before moving to the proof of the theorem, we prove that (R, C) and (R', C) are computationally indistinguishable.

Lemma 4.2. *Let $p : \{0,1\}^* \to \{0,1\}$ be a $(\sigma(n), \nu(n))$-hardcore predicate for π. Then, for R, R' and C as defined above it holds that:*

$$(R, C) \sim_{\varepsilon(n), s(n)} (R', C)\,,$$

where $\varepsilon(n) = n\nu(n)$ and $s(n) = \sigma(n) - \mathcal{O}(n(\sigma_p(n) + \sigma_\pi(n))$.

Proof. Assume that there exists an algorithm D running in $s(n)$ steps, for which

$$|\mathsf{Pr}\left[\mathsf{D}(R, C) = 1\right] - \mathsf{Pr}\left[\mathsf{D}(R', C) = 1\right]| > \varepsilon(n)\,.$$

Consider the following series of hybrids. The distribution of $\mathcal{H}_0$ is given by $(R', C_0) = (R', C)$. Now, when moving from $\mathcal{H}_i$ to $\mathcal{H}_{i+1}$, C is modified as follows: one element $\pi(x_j)$ of C_i satisfying $p(x_j) = b$, for which x_j is not part of R', is substituted by a random $\pi(\bar{x}_j)$ satisfying $p(\bar{x}_j) = 1 - b$, and C_{i+1} is reordered lexicographically.

Then, by definition, we have that $(R', C_0) = (R', C)$. Furthermore, it can be seen that over the random choices of $B \xleftarrow{\$} \{0,1\}$, it holds that $(R', C_n) = (R, C)$. Furthermore, there exists an i such D can distinguish (R', C_i) and (R', C_{i+1}) with advantage at least $\varepsilon(n)/n$.

We now show how D (outputting either i or $i + 1$ for simplicity) can be turned into an algorithm A of roughly the same running time, which predicts $p(x)$ given $\pi(x)$ for a uniformly chosen x with probability at least $\frac{1}{2} + \frac{\varepsilon(n)}{n}$. On input $y = \pi(x)$, A proceeds as follows:

– A uniformly guesses a bit $b' \xleftarrow{\$} \{0,1\}$;

- it then computes $x_1, \ldots, x_{2n-i-1} \xleftarrow{\$} \{0,1\}^n$ satisfying $p(x_j) = b'$, as well as $x_{2n+1-i}, \ldots, x_{3n} \xleftarrow{\$} \{0,1\}$ for which $p(x_j) = 1 - b'$;
- A then calls D on $\pi(x_1), \ldots, \pi(x_{2n-i-1}), y, \pi(x_{2n+1-i}), \ldots, \pi(x_{3n})$, sorted lexicographically;
- finally, A outputs b' if D returned i, and $1 - b'$ otherwise.

It can be seen that A's input to D is a sample of (R', C_i), if the secret $p(x) = b'$, and a sample of (R', C_{i+1}) otherwise for a random b'. It thus follows that A guesses $p(x)$ correctly with the same probability as D is able to distinguish (R', C_i) and (R', C_{i+1}) for random bit b. The complexity of A is essentially that of D, plus that for drawing, on average, $6n$ random elements in $\{0,1\}^n$ and evaluating π and p on those, yielding a contradiction to p being a $(\sigma(n), \nu(n))$-hardcore predicate. □

Proof (of Theorem 4.1). The claim is proved in two steps.

A Lower Bound for $H^{\mathsf{HILL}}_{\varepsilon,s}(R'|C)$. By Lemma 4.2 we have that $(R, C) \sim_{\varepsilon,s} (R', C)$. We thus get that

$$\begin{aligned} H^{\mathsf{HILL}}_{\varepsilon,s}(R'|C) \geq \widetilde{H}_\infty(R|C) &= -\log\left(\binom{2n}{n}\binom{n}{n/2}\right)^{-1} \\ &= \log\binom{2n}{n} + \log\binom{2\frac{n}{2}}{\frac{n}{2}} = 3n + \mathcal{O}(\log n)\,, \end{aligned}$$

where the first equality holds because R is uniformly distributed in its domain and $|R|$ does not depend on C, and the last one holds by (2). For sufficiently large n, this expression is lower bounded by $2.95n$.

An Upper Bound for $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'|C,B)$. Recap that $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'|C,B) \geq k$ if there exists a distribution X such that $(X,C,B) \sim_{\varepsilon',s'} (R',C,B)$, and $\widetilde{H}_\infty(X|C,B) \geq k$. To prove our theorem we will now prove an upper bound on $H^{\mathsf{HILL}}_{\varepsilon',s'}(R'|C,B)$ by showing that the conditional average min-entropy of every X satisfying $(X,C,B) \sim_{\varepsilon',s'} (R',C,B)$, is not significantly larger than the conditional average min-entropy of R'.

Let now X be such that the joint distribution (R',C,B) and (X,C,B) are close. We then observe that:

$$\mathsf{Pr}\left[X \notin \mathrm{supp}(R'(c,b)) \,:\, b \xleftarrow{\$} \{0,1\}, c \xleftarrow{\$} \mathsf{C}(1^n, b)\right] < \varepsilon'\,.$$

This holds because given (x,c,b), we can efficiently verify if $x \in \mathrm{supp}(R')$ or not: simply check that for exactly n components of x, their hardcore predicate evaluates to $1-b$, and secondly, that all components of x occur in c. Thus, if the probability X falling in the support of R' is more than ε', there exists an efficient distinguisher that tells the two distributions apart with advantage more than ε'.

Now, call a pair (c, b) *bad* if the above probability is larger than $\frac{1}{1.01}$, else, call it *good*. Then, by Markov's inequality, the fraction of bad (c, b) is at most $1.01\varepsilon'$. We then get that:

$$\begin{aligned}
\widetilde{H}_\infty(X|C,B) &= -\log \mathop{\mathbb{E}}_{c,b} \max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right] \\
&= -\log\left(\sum_{c,b} \mathsf{Pr}\left[C = c \wedge B = b\right] \max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right]\right) \\
&\leq -\log\left(\sum_{good\ (c,b)} \mathsf{Pr}\left[C = c \wedge B = b\right] \max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right]\right. \\
&\qquad \left. + \sum_{bad\ (c,b)} \mathsf{Pr}\left[C = c \wedge B = b\right] \max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right]\right) \\
&\leq -\log\left(\sum_{good\ (c,b)} \mathsf{Pr}\left[C = c \wedge B = b\right] \max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right]\right)
\end{aligned}$$

Using that for each (c, b), R' is uniformly distributed in its support, and that for good pairs we have that $\mathsf{Pr}\left[X \in \mathrm{supp}(R'(c,b))\right] > 1 - \frac{1}{1.01} = \frac{1}{101}$, we get that $\max_x \mathsf{Pr}\left[X = x | C = c \wedge B = b\right]$ is upper bounded by

$$\frac{1}{101} \max_r \mathsf{Pr}\left[R' = r | C = c \wedge B = b\right] = \frac{1}{101}\binom{2n}{n/2}^{-1},$$

which follows directly from the definition of R'. Using further that a fraction of at least $1 - 1.01\varepsilon'$ of all (b, c) is good, this now allows us to continue the above inequality chain by:

$$\begin{aligned}
&\leq -\log\left(\sum_{good(c,b)} \frac{\mathsf{Pr}\left[C = c \wedge B = b\right]}{101} \max_r \mathsf{Pr}\left[R' = r | C = c \wedge B = b\right]\right) \\
&\leq -\log\left((1 - 1.01\varepsilon')\frac{1}{101}\binom{2n}{n/2}^{-1}\right) \\
&= 4\frac{n}{2}\log 4 - 3\frac{n}{2}\log 3 + \mathcal{O}(\log n) - \log\left(\frac{1 - 1.01\varepsilon'}{101}\right) \\
&< 1.65n + \mathcal{O}(\log n) + 20,
\end{aligned}$$

where the last two inequality follow from (2) and our choice of ε'.

Now, for sufficiently large n, we get that this term is upper bounded by $1.7n$, and the claim of the theorem follows. □

5 Conclusion

Computational notions of entropy have found many applications in cryptography, and chain rules are a central tool in many security proofs. We showed that

the chain rule for one (arguably the most) important such notion, namely HILL entropy, does not hold.

Given that the chain rule holds and has been used for several variants (like relaxed, decomposable or simulatable) of HILL entropy, the question arises whether the current standard notion of conditional HILL entropy is the natural one to work with. We don't have an answer to this, but our results indicate that it is the only right notion in at least one natural setting, namely when talking about deniable encryption.

We hope the connection between chain rules for HILL entropy and deniable encryption we show will open new venues towards constructing the first deniable encryption scheme.

Acknowledgment. The authors want to thank Sasha Rubin for insightful comments and discussions while working on this paper.

References

1. Barak, B., Shaltiel, R., Wigderson, A.: Computational Analogues of Entropy. In: Arora, S., Jansen, K., Rolim, J.D.P., Sahai, A. (eds.) RANDOM 2003 and APPROX 2003. LNCS, vol. 2764, pp. 200–215. Springer, Heidelberg (2003)
2. Bendlin, R., Nielsen, J.B., Nordholt, P.S., Orlandi, C.: Lower and Upper Bounds for Deniable Public-Key Encryption. In: Lee, D.H., Wang, X. (eds.) ASIACRYPT 2011. LNCS, vol. 7073, pp. 125–142. Springer, Heidelberg (2011)
3. Canetti, R., Dwork, C., Naor, M., Ostrovsky, R.: Deniable Encryption. In: Kaliski Jr., B.S. (ed.) CRYPTO 1997. LNCS, vol. 1294, pp. 90–104. Springer, Heidelberg (1997)
4. Chung, K.-M., Kalai, Y.T., Liu, F.-H., Raz, R.: Memory Delegation. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 151–168. Springer, Heidelberg (2011)
5. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. SIAM Journal on Computing 38(1), 97–139 (2008)
6. Dürmuth, M., Freeman, D.M.: Deniable Encryption with Negligible Detection Probability: An Interactive Construction. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 610–626. Springer, Heidelberg (2011), Full version including a description of the flaw available at `http://eprint.iacr.org/2011/066.pdf`
7. Dziembowski, S., Pietrzak, K.: Leakage-Resilient Cryptography. In: FOCS 2008, pp. 293–302. IEEE Computer Society (2008)
8. Fuller, B., O'Neill, A., Reyzin, L.: A Unified Approach to Deterministic Encryption: New Constructions and a Connection to Computational Entropy. In: Cramer, R. (ed.) TCC 2012. LNCS, vol. 7194, pp. 582–599. Springer, Heidelberg (2012)
9. Fuller, B., Reyzin, L.: Computational Entropy and Information Leakage. Cryptology ePrint Archive, Report 2012/466 (2012), `http://eprint.iacr.org/`
10. Gentry, C., Wichs, D.: Separating Succinct Non-Interactive Arguments from All Falsifiable Assumptions. In: STOC 2011, pp. 99–108 (2011)
11. Goldreich, O.: Foundations of Cryptography: Basic Tools. Cambridge University Press, New York (2000)

12. Goldreich, O., Levin, L.A.: A Hard-Core Predicate for all One-Way Functions. In: Johnson, D.S. (ed.) STOC 1989, pp. 25–32. ACM (1989)
13. Håstad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A Pseudorandom Generator from any One-way Function. SIAM Journal on Computing 28(4), 1364–1396 (1999)
14. Hsiao, C.-Y., Lu, C.-J., Reyzin, L.: Conditional Computational Entropy, or Toward Separating Pseudoentropy from Compressibility. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 169–186. Springer, Heidelberg (2007)
15. Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S.P.: Dense Subsets of Pseudorandom Sets. In: FOCS 2008, pp. 76–85. IEEE Computer Society (2008)
16. Reyzin, L.: Some Notions of Entropy for Cryptography. In: Fehr, S. (ed.) ICITS 2011. LNCS, vol. 6673, pp. 138–142. Springer, Heidelberg (2011)
17. Shoup, V.: A Computational Introduction to Number Theory and Algebra. Cambridge Press (2009)
18. Trevisan, L.: Cryptography. Lecture Notes from CS 276 (2009)
19. Wee, H.M.: One-Way Permutations, Interactive Hashing and Statistically Hiding Commitments. In: Vadhan, S.P. (ed.) TCC 2007. LNCS, vol. 4392, pp. 419–433. Springer, Heidelberg (2007)
20. Yao, A.C.: Theory and Applications of Trapdoor Functions (Extended Abstract). In: FOCS 1982, pp. 80–91. IEEE Computer Society (1982)