Min-Entropy Leakage of Channels in Cascade

Barbara Espinoza and Geoffrey Smith

School of Computing and Information Sciences
Florida International University, Miami FL 33199, USA
{bespi009,smithg}@cis.fiu.edu

Abstract. Theories of quantitative information flow offer an attractive framework for analyzing confidentiality in practical systems, which often cannot avoid "small" leaks of confidential information. Recently there has been growing interest in the theory of min-entropy leakage, which measures uncertainty based on a random variable's vulnerability to being guessed in one try by an adversary. Here we contribute to this theory by studying the min-entropy leakage of systems formed by cascading two channels together, using the output of the first channel as the input to the second channel. After considering the semantics of cascading carefully and exposing some technical subtleties, we prove that the min-entropy leakage of a cascade of two channels cannot exceed the leakage of the first channel; this result is a min-entropy analogue of the classic data-processing inequality. We show however that a comparable bound does not hold for the second channel. We then consider the min-capacity, or maximum leakage over all *a priori* distributions, showing that the min-capacity of a cascade of two channels cannot exceed the min-capacity of either channel.

1 Introduction

Protecting confidential information from improper disclosure is a fundamental security goal, made more challenging by the unavoidability of "small" information leaks in practical systems. In the past decade, there has been growing interest in *quantitative* theories of information flow [7,15] that allow us to talk about "how much" information is leaked and (perhaps) allow us to tolerate "small" leaks. One theory of quantitative information flow that has received considerable attention recently [21,6,4,13,3] is based on measuring uncertainty using Rényi's *min-entropy* [18], rather than using Shannon entropy [20]. The advantage of min-entropy leakage is that it is based directly on a secret's *vulnerability* to being guessed in one try by an adversary, resulting in stronger operational security guarantees than are obtained with Shannon entropy and mutual information [21]. (We review the theory of min-entropy leakage in Section 3.)

The basis for all information-theoretic measures of information flow is the concept of a *channel*, which consists of a set $\mathcal{S}$ of secret inputs, a set $\mathcal{O}$ of observable outputs, and a *channel matrix* $C_{\mathcal{SO}}$, which specifies the conditional probability of obtaining output $o \in \mathcal{O}$, given input $s \in S$. Given an *a priori* distribution on $\mathcal{S}$, a measure of information flow specifies how much information flows from random variable S to random variable O. Moreover, the *capacity* of $C_{\mathcal{SO}}$ is the maximum amount of flow over all *a priori* distributions on $\mathcal{S}$.

G. Barthe, A. Datta, and S. Etalle (Eds.): FAST 2011, LNCS 7140, pp. 70–84, 2012.

A classic construction on two channels is *cascading* [9,1], where the output of the first channel is used as the input to the second. A natural question concerns the amount of information flow in a cascade of channels, as compared with in each of the two channels. In the theory of mutual-information flow, the classic *data-processing inequality* [8, p. 34] says that the mutual-information flow on a cascade of channels cannot exceed the flow on either channel; this straightforwardly implies similar bounds for Shannon capacity. In this paper, our main goal is to investigate whether similar properties hold for min-entropy leakage.

A bound on the min-capacity of a cascade of channels was shown earlier by Köpf and Smith [16]. They showed that if a channel $C_{\mathcal{SO}}$ can be factored into the cascade of channels $C_{\mathcal{ST}}$ and $C_{\mathcal{TO}}$, then the min-capacity of $C_{\mathcal{SO}}$ is at most the logarithm of $|\mathcal{T}|$, the size of the set of intermediate results. They used this result to establish security guarantees of blinded cryptography under timing attacks, modeling such an attack as a channel whose input is a secret decryption key and whose output is a sequence of timings of decryption operations using that key. They showed that this channel can be factored into the cascade of two channels such that the set of intermediate results is small, which implies that its min-capacity is small.

Our main contribution in this paper is to go beyond the results of [16] by establishing results not just on min-capacity, but also on min-entropy leakage under a given *a priori* distribution. In particular, we show that under any *a priori* distribution, the min-entropy leakage of a cascade of channels cannot exceed the leakage of the first link, and show that, contrary to our intuition, it *can* exceed the leakage of the second link. Given the cascade of $C_{\mathcal{ST}}$ and $C_{\mathcal{TO}}$, we also compare the conditional vulnerabilities $V(S|O)$, $V(S|T)$, and $V(T|O)$.[1] We show that $V(S|O) \leq V(S|T)$, but that no relationship need hold between $V(S|O)$ and $V(T|O)$. In the case when $C_{\mathcal{ST}}$ is deterministic, however, we show that $V(S|O) \leq V(T|O)$. Turning to min-capacity, we generalize the results of [16], showing that the min-capacity of a cascade of channels is upper bounded not just by the logarithm of the number of intermediate results, but also by the min-capacity of each of the links. These results give us a general technique for bounding the min-entropy leakage of any channel that can be factored into a cascade of channels.

An additional contribution of this paper is that we study carefully the semantics of cascading of channels, exposing some technical subtleties with non-uniqueness of joint distributions and also taking care to deal with undefined conditional probabilities.

The rest of the paper is structured as follows. In Section 2 we review the notions of channel and cascade of channels, carefully dealing with undefined conditional probabilities, and pointing out some nuances in the standard definitions. In Section 3 we present a review of the min-entropy measure of information flow. In Section 4 we present our results on the min-entropy leakage and vulnerability of a cascade of channels under a given *a priori* distribution. In Section 5 we extend these leakage results to results on the min-capacity of a cascade of channels. Finally, in Sections 6 and 7 we discuss related work and conclude.

[1] As will be reviewed in Section 3, $V(S|O)$ is the expected probability of guessing the value of S, given the value of O.

2 Foundations of Channels and Cascades of Channels

2.1 Channels

A *channel* is a triple $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$, where $\mathcal{S}$ is a finite set of secret input values, $\mathcal{O}$ is a finite set of observable output values, and $C_{\mathcal{SO}}$ is a $|\mathcal{S}| \times |\mathcal{O}|$ matrix, called the *channel matrix*, such that $C_{\mathcal{SO}}[s, o]$ is the *conditional probability* of obtaining output o when the input is s. Note that each entry of $C_{\mathcal{SO}}$ is between 0 and 1, and each row sums to 1:

$$\text{for every } s \in \mathcal{S}, \sum_{o \in \mathcal{O}} C_{\mathcal{SO}}[s, o] = 1. \tag{1}$$

An important special case is a *deterministic channel*, where each input yields a unique output. In terms of $C_{\mathcal{SO}}$, this means that each entry is either 0 or 1, and each row contains exactly one 1.

Recall that in traditional probability theory, conditional probabilities are defined in terms of joint distributions. So, in the absence of a joint distribution, how can we speak of $C_{\mathcal{SO}}$ as giving conditional probabilities? We believe that it is actually best to view these conditional probabilities as a *primitive notion*—they simply say that *if* the input is s, *then* output o will occur with probability $C_{\mathcal{SO}}[s, o]$.[2]

We are interested in studying the behavior of a channel $C_{\mathcal{SO}}$ under an *a priori* distribution P_S on $\mathcal{S}$, which gives a random variable S. Now we can show that there is a unique joint distribution $P^\star_{SO}$ on $\mathcal{S} \times \mathcal{O}$ such that

1. $P^\star_{SO}$ recovers the *a priori* P_S by marginalization, and
2. $P^\star_{SO}$ recovers the conditional probabilities $C_{\mathcal{SO}}$, whenever they are defined.

To see that there is at most one such joint distribution, note first that if $P_S[s] = 0$, then by condition 1 we must have $0 = P^\star_S[s] = \sum_{o \in \mathcal{O}} P^\star_{SO}[s, o]$, which implies that $P^\star[s, o] = 0$, for every $o \in \mathcal{O}$. Second, if $P_S[s] \neq 0$, then by conditions 1 and 2 we must have, for every $o \in \mathcal{O}$,

$$C_{\mathcal{SO}}[s, o] = P^\star_{O|S}[o|s] = \frac{P^\star_{SO}[s, o]}{P^\star_S[s]} = \frac{P^\star_{SO}[s, o]}{P_S[s]}$$

which implies that $P^\star_{SO}[s, o] = P_S[s] C_{\mathcal{SO}}[s, o]$. Finally, observe that these two cases can be merged into a single definition:

$$P^\star_{SO}[s, o] = P_S[s] C_{\mathcal{SO}}[s, o]. \tag{2}$$

Equivalently, we can define $P^\star_{SO}$ as the product of a diagonal matrix with P_S on its diagonal, and $C_{\mathcal{SO}}$:

$$P^\star_{SO} = \text{diag}(P_S) C_{\mathcal{SO}}. \tag{3}$$

Now we show that $P^\star_{SO}$ indeed has the properties that we want:

Theorem 1. *$P^\star_{SO}$ is the unique joint distribution that recovers P_S by marginalization and recovers the conditional probabilities $C_{\mathcal{SO}}$, whenever they are defined.*

[2] Indeed, Rényi argued that "the basic notion of probability theory should be the notion of the conditional probability of A under the condition B" [19, p. 35].

Proof. We have already argued the uniqueness of $P^\star_{SO}$. Now, $P^\star_{SO}$ recovers P_S by marginalization, since for any $s \in \mathcal{S}$,

$$P^\star_S[s] = \sum_{o \in \mathcal{O}} P^\star_{SO}[s,o] = \sum_{o \in \mathcal{O}} P_S[s] C_{\mathcal{SO}}[s,o] = P_S[s] \sum_{o \in \mathcal{O}} C_{\mathcal{SO}}[s,o] = P_S[s].$$

From this, we also see that $P^\star_{SO}$ is a valid distribution, since

$$\sum_{s \in \mathcal{S}, o \in \mathcal{O}} P^\star_{SO}[s,o] = \sum_{s \in \mathcal{S}} \sum_{o \in \mathcal{O}} P^\star_{SO}[s,o] = \sum_{s \in \mathcal{S}} P_S[s] = 1.$$

Finally, $P^\star_{SO}$ recovers the conditional probabilities $C_{\mathcal{SO}}$, whenever they are defined. For if $P_S[s] \neq 0$, then for any $o \in \mathcal{O}$,

$$P^\star_{O|S}[o|s] = \frac{P^\star_{SO}[s,o]}{P^\star_S[s]} = \frac{P_S[s] C_{\mathcal{SO}}[s,o]}{P_S[s]} = C_{\mathcal{SO}}[s,o].$$

□

We also get a distribution on $\mathcal{O}$ by marginalization, giving a random variable O:

$$P^\star_O[o] = \sum_{s \in \mathcal{S}} P^\star_{SO}[s,o].$$

Following Gallager [12], we will omit the subscripts from probability distributions whenever they are clear from context, for example writing $P^\star[s|o]$ instead of $P^\star_{S|O}[s|o]$. Also, with a slight abuse of notation, we will sometimes use $P^\star_{O|S}$ to denote the channel matrix from $\mathcal{S}$ to $\mathcal{O}$ whose entries are the conditional probabilities recovered from $P^\star$.

2.2 Cascades of Channels

Given channels $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ and $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$, where the set of outputs of the first is the same as the set of inputs of the second, it makes sense to form a *cascade of channels* that composes the channels sequentially [1]. Intuitively, given an *a priori* distribution P_S, the cascade of channels will proceed in two steps. First, the information in S flows through the first channel and determines a distribution P_T and a random variable T. Then, the information in T flows through the second channel to produce the final output distributed according to P_O.

When we consider the formal semantics of a cascade of channels, we might expect (based on the previous section) that there is a unique joint distribution P_{STO} that recovers P_S and the conditional probabilities $C_{\mathcal{ST}}$ and $C_{\mathcal{TO}}$, whenever they are defined. Curiously, this turns out not to be true.

Example 1. Let $\mathcal{S} = \mathcal{T} = \mathcal{O} = \{0,1\}$, and let $C_{\mathcal{ST}}$, $C_{\mathcal{TO}}$, and P_S be as follows:

$$C_{\mathcal{ST}} = \begin{pmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix} \qquad C_{\mathcal{TO}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \qquad P_S = (2/3, 1/3).$$

With this setup, we can pinpoint at least two scenarios for the joint distribution P_{STO}. Recall that any joint distribution must satisfy the product rule

$$P[s,t,o] = P[s]P[t|s]P[o|s,t]$$

whenever the conditional probabilities are defined. Since we demand $P[s] = P_S[s]$ and $P[t|s] = C_{\mathcal{ST}}[s,t]$, it is clear that our only freedom is in choosing $P[o|s,t]$.

For our first scenario, we make O the *exclusive or* of S and T:

$$P^{\oplus}[o|s,t] = \begin{cases} 1, & \text{if } o = s \oplus t \\ 0, & \text{otherwise} \end{cases}$$

Using the product rule, we obtain the following joint distribution:

S	T	O	$P^{\oplus}[s,t,o]$
0	0	0	$1/6$
0	0	1	0
0	1	0	0
0	1	1	$1/2$
1	0	0	0
1	0	1	$1/6$
1	1	0	$1/6$
1	1	1	0

This joint distribution $P^{\oplus}$ recovers P_S as well as the conditional probabilities $C_{\mathcal{ST}}$ and $C_{\mathcal{TO}}$. For example, we can verify that $P^{\oplus}_{O|T}[0|1] = 1/4 = C_{\mathcal{TO}}[1,0]$:

$$P^{\oplus}_{O|T}[0|1] = \frac{P^{\oplus}_{TO}[1,0]}{P^{\oplus}_{T}[1]} = \frac{\sum_s P^{\oplus}_{STO}[s,1,0]}{\sum_{s,o} P^{\oplus}_{STO}[s,1,o]} = \frac{0 + 1/6}{0 + 1/2 + 1/6 + 0} = 1/4.$$

Note, however, the definition of $P^{\oplus}$ is contrary to our intended "cascading" behavior, since it makes the conditional probability of O depend on *both* S and T.[3]

For our second scenario, we instead make the conditional probability of O depend only on T, choosing $P^{\star}[o|s,t] = P^{\star}[o|t]$. This gives a second joint distribution that recovers P_S and the conditional probabilities $C_{\mathcal{ST}}$ and $C_{\mathcal{TO}}$:

S	T	O	$P^{\star}[s,t,o]$
0	0	0	$1/12$
0	0	1	$1/12$
0	1	0	$1/8$
0	1	1	$3/8$
1	0	0	$1/12$
1	0	1	$1/12$
1	1	0	$1/24$
1	1	1	$1/8$

□

[3] A strange consequence is that $P^{\oplus}_{O|T}$ depends on the *a priori* P_S. For instance, if we change P_S to $(1/2, 1/2)$, we find that $P^{\oplus}_{O|T}$ no longer coincides with $C_{\mathcal{TO}}$.

Using the intuitions developed in Example 1, we formally define the semantics of a cascade of channels:

Definition 1. *The* cascade *of channels* $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ *and* $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$ *under* a priori *distribution* P_S *has joint distribution* $P^\star_{STO}$*, where*

$$P^\star_{STO}[s,t,o] = P_S[s]C_{\mathcal{ST}}[s,t]C_{\mathcal{TO}}[t,o].$$

We now establish the properties of $P^\star_{STO}$ in a series of theorems, whose proofs are similar to that of Theorem 1. Due to space limitations, the proofs are omitted.

Theorem 2. $P^\star_{STO}$ *recovers the* a priori P_S *by marginalization.*

Theorem 3. $P^\star_{STO}$ *is a valid joint distribution.*

Theorem 4. $P^\star_{STO}$ *recovers the conditional probabilities* $C_{\mathcal{ST}}$ *and* $C_{\mathcal{TO}}$*, whenever they are defined.*

Theorem 5. *Whenever* $P^\star[s,t] \neq 0$*, we have* $P^\star[o|s,t] = P^\star[o|t]$.

Moreover, $P^\star$ is the *unique* joint distribution that satisfies these four theorems:

Theorem 6. *If* P_{STO} *is any joint distribution that recovers* P_S*, gives the correct conditional probabilities when they are defined, and satisfies* $P[o|s,t] = P[o|t]$ *when they are defined,* P_{STO} *is equal to* $P^\star_{STO}$.

We next turn our attention to the conditional probabilities $P^\star[o|s]$, showing that these can be obtained by matrix multiplication:

Theorem 7. *Whenever* $P_S[s] \neq 0$*, we have* $P^\star[o|s] = C_{\mathcal{ST}}C_{\mathcal{TO}}[s,o]$.

This last property motivates the following definition, which specifies the cascade of channels independently of an *a priori* distribution:

Definition 2. *The* cascade *of channels* $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ *and* $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$ *is the channel* $(\mathcal{S}, \mathcal{O}, C_{\mathcal{ST}}C_{\mathcal{TO}})$.

Remark 1. Recalling Example 1, we can calculate that

$$P^\star_{O|S} = \begin{pmatrix} 5/16 & 11/16 \\ 3/8 & 5/8 \end{pmatrix} = \begin{pmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} = C_{\mathcal{ST}}C_{\mathcal{TO}}.$$

In contrast,

$$P^\oplus_{O|S} = \begin{pmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix}.$$

This might make us wonder whether the property that $P_{O|S}$ is given by matrix multiplication might suffice to determine $P^\star_{STO}$. But this turns out not to be true. Consider the channels

$$C_{\mathcal{ST}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \qquad C_{\mathcal{TO}} = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix} \qquad P_S = (2/3, 1/3).$$

If we define $P^\oplus_{STO}$ as in Example 1, then we get $P^\oplus_{STO} \neq P^\star_{STO}$, but nevertheless

$$P^\oplus_{O|S} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = C_{\mathcal{ST}}C_{\mathcal{TO}}.$$

□

2.3 Factoring Channels

Suppose that we can *factor* a channel matrix $C_{\mathcal{SO}}$ into a product of two channel matrices, so that $C_{\mathcal{SO}} = C_{\mathcal{ST}}C_{\mathcal{TO}}$. This procedure introduces a new set of intermediate values $\mathcal{T}$, giving two channels $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ and $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$ whose cascade is the original channel. When dealing with channel matrix factorizations, we will refer to the number of elements in $\mathcal{T}$ as the *inner dimension* of the factorization.

3 Measuring Information Flow Using Min-Entropy

Given a channel $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$, we consider an adversary $\mathcal{A}$ that wishes to guess the value of S. We assume that $\mathcal{A}$ knows both the *a priori* distribution P_S and the channel. It is then natural to measure the amount of information that flows from S to O by considering the reduction in $\mathcal{A}$'s uncertainty about S after observing the value of O.

$$\text{leakage} = \text{initial uncertainty} - \text{remaining uncertainty}. \tag{4}$$

We define uncertainty in terms of worst-case probability that $\mathcal{A}$ will guess the correct value of S in one try. This measure is known as the *vulnerability* of S and has been defined in [21]. We distinguish between the vulnerability before and after observing the value of O. The former is called the *a priori* vulnerability and defined as

$$V(S) = \max_{s \in \mathcal{S}} P_S[s].$$

The latter is the *a posteriori* vulnerability and is defined as the expected vulnerability after observing the value of O.

$$\begin{aligned} V(S|O) &= \sum_{o \in \mathcal{O}} P^\star[o] \max_{s \in \mathcal{S}} P^\star[s|o] \\ &= \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} P^\star[s, o] \\ &= \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} (P_S[s] C_{\mathcal{SO}}[s, o]). \end{aligned}$$

We can convert from probability measures to bit measures by taking the negative logarithm. Using this method, we obtain our measures of uncertainty.

- initial uncertainty: $H_\infty(S) = -\log V(S)$.
- remaining uncertainty: $H_\infty(S|O) = -\log V(S|O)$.

In information theory, the quantity H_∞ is known as Rényi *min-entropy*. The notation $H_\infty(S|O)$ should then be read as the conditional min-entropy of S given O.

Substituting our uncertainty measures in equation (4) we can define the *min-entropy leakage* from S to O, denoted by $\mathcal{L}_{SO}$, to be

$$\mathcal{L}_{SO} = H_\infty(S) - H_\infty(S|O) = -\log V(S) - (-\log V(S|O)) = \log \frac{V(S|O)}{V(S)}.$$

Thus, the min-entropy leakage is the logarithm of the factor by which knowledge of O increases the vulnerability of S.

An important notion in information theory is the *channel capacity*, which is the maximum leakage over all possible *a priori* distributions. In the case of min-entropy leakage, we will refer to this measure as the *min-capacity* of the channel and use the notation $\mathcal{ML}(C_{\mathcal{SO}})$:

$$\mathcal{ML}(C_{\mathcal{SO}}) = \sup_{P_S \in \mathcal{D}(\mathcal{S})} \mathcal{L}_{SO}.$$

The min-capacity is always realized by a uniform distribution on S (and possibly by other distributions as well) [6,16], and can be easily calculated as the logarithm of the sum of the column maximums in $C_{\mathcal{SO}}$.

$$\mathcal{ML}(C_{\mathcal{SO}}) = \log \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} C_{\mathcal{SO}}[s, o].$$

As a consequence, the min-capacity of $C_{\mathcal{SO}}$ is 0 iff $C_{\mathcal{SO}}$ has no leakage at all, in that all of its rows are identical [16].

4 Leakage in a Cascade of Channels

In this section we explore how the min-entropy leakage behaves in a cascade of channels by comparing the leakage of each of the links with the total leakage.

If we imagine channels as pipes, and information as water that flows through these pipes, then we might anticipate that the leakage in a cascade of channels cannot exceed the leakage of the first link. We prove this property in Theorem 8.

Theorem 8. *Let $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ be the cascade of $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ and $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$. Then for any* a priori *distribution P_S, we have $\mathcal{L}_{SO} \leq \mathcal{L}_{ST}$.*

Proof. Unfolding the formula of min-entropy leakage, we observe that the desired inequality is equivalent to an inequality on the conditional vulnerabilities:

$$\mathcal{L}_{SO} \leq \mathcal{L}_{ST} \iff \log \frac{V(S|O)}{V(S)} \leq \log \frac{V(S|T)}{V(S)} \iff V(S|O) \leq V(S|T).$$

Those conditional vulnerabilities are the sum of the column maximums in the corresponding joint matrices:

$$V(S|O) = \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} P^{\star}_{SO}[s, o] \qquad V(S|T) = \sum_{t \in \mathcal{T}} \max_{s \in \mathcal{S}} P^{\star}_{ST}[s, t].$$

Recall from equation (3) that we can express the joint matrices as a matrix product:

$$P^{\star}_{SO} = \text{diag}(P_S) C_{\mathcal{SO}} \qquad P^{\star}_{ST} = \text{diag}(P_S) C_{\mathcal{ST}}.$$

Considering that $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ is a cascade of channels we get

$$P^{\star}_{SO} = \text{diag}(P_S) C_{\mathcal{SO}} = \text{diag}(P_S)(C_{\mathcal{ST}} C_{\mathcal{TO}}) = (\text{diag}(P_S) C_{\mathcal{ST}}) C_{\mathcal{TO}} = P^{\star}_{ST} C_{\mathcal{TO}}.$$

Hence, it is our goal to prove that the sum of the column maximums in $P^\star_{ST}$ must be at least as large as the sum of the column maximums in $P^\star_{ST}C_{\mathcal{TO}}$.[4]

Let α_t for $t \in \mathcal{T}$ denote the maximum of column t of $P^\star_{ST}$:

$$\alpha_t = \max_{s \in \mathcal{S}} P^\star_{ST}[s,t].$$

Also, let β_o denote the maximum of column o of $P^\star_{SO}$:

$$\beta_o = \max_{s \in \mathcal{S}} P^\star_{SO}[s,o].$$

Then, for every $o \in \mathcal{O}$, the elements in column o of P_{SO} satisfy

$$P^\star_{SO}[s,o] = \sum_{t \in \mathcal{T}} P^\star_{ST}[s,t]C_{\mathcal{TO}}[t,o] \leq \sum_{t \in \mathcal{T}} \alpha_t C_{\mathcal{TO}}[t,o].$$

In particular, this property is satisfied by the column maximum:

$$\beta_o \leq \sum_{t \in \mathcal{T}} \alpha_t C_{\mathcal{TO}}[t,o].$$

Then, using these properties we proceed with the proof:

$$\begin{aligned}
V(S|O) &= \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} P^\star_{SO}[s,o] \\
&= \sum_{o \in \mathcal{O}} \beta_o \\
&\leq \sum_{o \in \mathcal{O}} \sum_{t \in \mathcal{T}} \alpha_t C_{\mathcal{TO}}[t,o] \\
&= \sum_{t \in \mathcal{T}} \sum_{o \in \mathcal{O}} \alpha_t C_{\mathcal{TO}}[t,o] \\
&= \sum_{t \in \mathcal{T}} \alpha_t \sum_{o \in \mathcal{O}} C_{\mathcal{TO}}[t,o] \\
&= \sum_{t \in \mathcal{T}} \alpha_t \\
&= \sum_{t \in \mathcal{T}} \max_{s \in \mathcal{S}} P^\star_{ST}[s,t] \\
&= V(S|T).
\end{aligned}$$

□

Note that Theorem 8 can be understood as the min-entropy analogue to the classic *data-processing inequality* of information theory. The data-processing inequality can be read as saying that "data processing can only destroy information" [17, p. 141].

[4] Notice that the number of columns in $P^\star_{ST}$ and $P^\star_{SO}$ need not match, so the task cannot be reduced to comparing the matrices column by column.

The standard formulation of the data-processing inequality [8, p. 34] starts with the hypothesis that S, T, O form a *Markov chain*, denoted $S \to T \to O$, which means that the joint distribution satisfies the equality

$$P[s,t,o] = P_S[s]P[t|s]P[o|t]. \tag{5}$$

It says then that the flow from S to O cannot exceed the flow from S to T, as measured by mutual information:

$$I(S;O) \leq I(S;T).$$

A drawback of this formulation is that the *a priori* P_S is "hard coded" into the Markov chain, rather than being a separate parameter as in the formulation of Theorem 8. Moreover, equation (5) runs into undefined conditional probabilities if some values of $\mathcal{S}$ or $\mathcal{T}$ have probability 0.

However, we can observe that if we have a cascade of channels and an *a priori* distribution such that every value in $\mathcal{S}$ and $\mathcal{T}$ has non-zero probability, then there is a Markov chain $S \to T \to O$. Hence we can get a version of the data-processing inequality with a formulation similar to that of Theorem 8:

Theorem 9. *Let $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ be the cascade of $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ and $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$, and let P_S be an* a priori *distribution. If every value in $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{O}$ has non-zero probability,*[5] *then $I(S;O) \leq I(S;T)$.*

Returning now to min-entropy leakage, when we consider the leakage in the *second* link of a cascade of channels, we find that it does not behave in the same way as the leakage in the first link. In fact, as the following example shows, the leakage of a cascade of channels may exceed the leakage of the second link.

Example 2. Let P_S be a uniform distribution, and

$$C_{\mathcal{ST}} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad\qquad C_{\mathcal{TO}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, $P_T = (7/8, 1/8)$, and the leakage from S to O exceeds the leakage from T to O:

$$\mathcal{L}_{SO} = \log 2 \qquad\qquad \mathcal{L}_{TO} = \log 8/7.$$

To understand why, recall that the min-entropy leakage is the logarithm of the factor by which the vulnerability increases after observing the output. In this example, $V(S) = 1/8$ and $V(S|O) = 1/4$, so channel $C_{\mathcal{SO}}$ doubles the vulnerability of S, giving

[5] The assumption about the values in $\mathcal{O}$ is not stated explicitly in [8], but it is used implicitly in the proof there.

min-entropy leakage of $\log 2$. Now, channel $C_{\mathcal{TO}}$ is a *noiseless* channel, so it leaks T completely, giving $V(T|O) = 1$. But $V(T)$, the *a priori* vulnerability of T, is $7/8$, so channel $C_{\mathcal{TO}}$ cannot possibly increase the vulnerability of T by more than a factor of $\frac{1}{7/8} = 8/7$. □

Notice that when we compare $\mathcal{L}_{SO}$ and $\mathcal{L}_{ST}$ in Theorem 8, we are comparing $\log \frac{V(S|O)}{V(S)}$ and $\log \frac{V(S|T)}{V(S)}$, which then amounts to a comparison between $V(S|O)$ and $V(S|T)$. But when we compare $\mathcal{L}_{SO}$ and $\mathcal{L}_{TO}$, we are comparing $\log \frac{V(S|O)}{V(S)}$ and $\log \frac{V(T|O)}{V(T)}$, which means that the comparison depends on both the numerators and also the denominators.

Exploring further, we found that it is not even possible to establish that $V(S|O) \leq V(T|O)$ in general. As an example, if there is only one value of S and multiple possible values of T, then $V(S|O)$ is certainly equal to 1, while $V(T|O)$ could be less than 1.

However, given the additional assumption that $C_{\mathcal{ST}}$ is deterministic, we would expect that $V(S|O) \leq V(T|O)$. Intuitively, if we correctly guess S, then we can use $C_{\mathcal{ST}}$ to deduce T as well. We prove this in the following theorem:

Theorem 10. *If $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ is the cascade of $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ and $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$, where $C_{\mathcal{ST}}$ is deterministic, then for any* a priori *P_S we have $V(S|O) \leq V(T|O)$.*

Proof. Let $f : \mathcal{S} \to \mathcal{T}$ denote the function described by the deterministic channel $C_{\mathcal{ST}}$, that is, $f(s) = t \iff C_{\mathcal{ST}}[s, t] = 1$. Also, let $[s]_f$ be the set of elements in $\mathcal{S}$ that map to $f(s)$, that is, $[s]_f = \{s' \in \mathcal{S} \mid f(s') = f(s)\}$. Since $C_{\mathcal{ST}}$ is deterministic, for each $s \in \mathcal{S}$ the probability $P_S[s]$ is at most the probability of its image $P^\star_T[f(s)]$:

$$\begin{aligned}
P_S[s] &\leq \sum_{s' \in [s]_f} P_S[s'] \\
&= \sum_{s' \in [s]_f} P_S[s'] C_{\mathcal{ST}}[s', f(s)] \\
&= \sum_{s' \in [s]_f} P_S[s'] C_{\mathcal{ST}}[s', f(s)] + \sum_{s'' \in \mathcal{S} \setminus [s]_f} P_S[s''] C_{\mathcal{ST}}[s'', f(s)] \\
&= \sum_{s' \in \mathcal{S}} P_S[s'] C_{\mathcal{ST}}[s', f(s)] \\
&= P^\star_T[f(s)].
\end{aligned}$$

Furthermore, we can see that $C_{\mathcal{SO}}[s, o] = C_{\mathcal{TO}}[f(s), o]$:

$$\begin{aligned}
C_{\mathcal{SO}}[s, o] &= \sum_{t \in \mathcal{T}} C_{\mathcal{ST}}[s, t] C_{\mathcal{TO}}[t, o] \\
&= C_{\mathcal{ST}}[s, f(s)] C_{\mathcal{TO}}[f(s), o] + \sum_{t \in \mathcal{T} \setminus \{f(s)\}} C_{\mathcal{ST}}[s, t] C_{\mathcal{TO}}[t, o] \\
&= C_{\mathcal{ST}}[s, f(s)] C_{\mathcal{TO}}[f(s), o] \\
&= C_{\mathcal{TO}}[f(s), o].
\end{aligned}$$

Then, using the previous two properties we can proceed with the proof:

$$\begin{aligned} V(S|O) &= \sum_{o\in\mathcal{O}} \max_{s\in\mathcal{S}} (P_S[s] C_{\mathcal{SO}}[s,o]) \\ &= \sum_{o\in\mathcal{O}} \max_{s\in\mathcal{S}} (P_S[s] C_{\mathcal{TO}}[f(s),o]) \\ &\leq \sum_{o\in\mathcal{O}} \max_{s\in\mathcal{S}} (P_T^\star[f(s)] C_{\mathcal{TO}}[f(s),o]) \\ &= \sum_{o\in\mathcal{O}} \max_{t\in\mathcal{T}} (P_T^\star[t] C_{\mathcal{TO}}[t,o]) \\ &= V(T|O). \end{aligned}$$

□

Contrary to our results for min-entropy leakage, with Shannon mutual information leakage we get bounds on *both* links of the cascade [1]. We can easily prove the bound on the second link if we consider that a Markov chain $S \to T \to O$ implies another Markov chain $O \to T \to S$. So, by the data-processing inequality, we have $I(O;S) \leq I(O;T)$. But now we can use the *symmetry* of mutual information (i.e. the fact that $I(X;Y) = I(Y;X)$) to deduce that $I(S;O) \leq I(T;O)$.

Remark 2. The symmetry of mutual information is key in proving the data-processing inequality for the second link of a cascade. But it is arguably a strange property; it seems counterintuitive that the mutual information leakage from S to O should be the same as the mutual information leakage from O to S. Min-entropy leakage, in contrast, is not symmetric in general. As an example, consider the following $n \times (n+1)$ channel matrix:

$$C_{\mathcal{SO}} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & \dots & 0 \\ 1/2 & 0 & 1/2 & 0 & \dots & 0 \\ 1/2 & 0 & 0 & 1/2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1/2 & 0 & 0 & 0 & \dots & 1/2 \end{pmatrix}.$$

Under a uniform *a priori* distribution, $V(S) = \frac{1}{n}$ and $V(S|O) = \frac{n+1}{2n}$, which implies that $\mathcal{L}_{SO} = \log \frac{n+1}{2}$. But when we view $P^\star_{SO}$ as a channel from O to S, we find that $V(O) = \frac{1}{2}$ but also $V(O|S) = \sum_s \max_o P^\star_{SO}[s,o] = \frac{1}{2}$, which implies that $\mathcal{L}_{OS} = 0$.

□

5 Capacity of a Cascade of Channels

We can extend the result from Theorem 8 to the capacity of a cascade of channels.

Corollary 1. *If $(\mathcal{S},\mathcal{O},C_{\mathcal{SO}})$ is the cascade of $(\mathcal{S},\mathcal{T},C_{\mathcal{ST}})$ and $(\mathcal{T},\mathcal{O},C_{\mathcal{TO}})$, then $\mathcal{ML}(C_{\mathcal{SO}}) \leq \mathcal{ML}(C_{\mathcal{ST}})$.*

Proof. From Theorem 8 we know that for any *a priori* P_S, $\mathcal{L}_{SO} \leq \mathcal{L}_{ST}$. So, since $\mathcal{ML}(C_{\mathcal{ST}}) = \sup_{P_S \in \mathcal{D}_\mathcal{S}} \mathcal{L}_{ST}$, we have for any *a priori* P_S that

$$\mathcal{ML}(C_{\mathcal{ST}}) \geq \mathcal{L}_{ST} \geq \mathcal{L}_{SO},$$

so

$$\mathcal{ML}(C_{\mathcal{ST}}) \geq \sup_{P_S \in \mathcal{D}_\mathcal{S}} \mathcal{L}_{SO} = \mathcal{ML}(C_{\mathcal{SO}}).$$

□

We can also provide an alternative proof for the upper bound on the capacity of a cascade of channels from [16]. That is, the capacity of a cascade of channels cannot exceed the logarithm of number of intermediate results:

Corollary 2. *If* $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ *is the cascade of* $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ *and* $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$*, then* $\mathcal{ML}(C_{\mathcal{SO}}) \leq \log |\mathcal{T}|$.

Proof. We have $\mathcal{ML}(C_{\mathcal{SO}}) \leq \mathcal{ML}(C_{\mathcal{ST}})$. But $\mathcal{ML}(C_{\mathcal{ST}})$ is the logarithm of the sum of the column maximums of $C_{\mathcal{ST}}$. Since $C_{\mathcal{ST}}$ has $|\mathcal{T}|$ columns, and each maximum is at most 1, we have $\mathcal{ML}(C_{\mathcal{SO}}) \leq \log |\mathcal{T}|$. □

Finally, unlike our result for min-entropy leakage under *a priori* P_S, we can prove that the min-capacity of a cascade of channels cannot exceed the min-capacity of the second link.

Theorem 11. *If* $(\mathcal{S}, \mathcal{O}, C_{\mathcal{SO}})$ *is the cascade of* $(\mathcal{S}, \mathcal{T}, C_{\mathcal{ST}})$ *and* $(\mathcal{T}, \mathcal{O}, C_{\mathcal{TO}})$*, then* $\mathcal{ML}(C_{\mathcal{SO}}) \leq \mathcal{ML}(C_{\mathcal{TO}})$.

Proof. The rows of $C_{\mathcal{SO}}$ are a convex combination of the rows of $C_{\mathcal{TO}}$. Hence, for each $o \in \mathcal{O}$, the elements in column o of $C_{\mathcal{SO}}$ are at most the maximum of column o of $C_{\mathcal{TO}}$:

$$C_{\mathcal{SO}}[s, o] = C_{\mathcal{ST}} C_{\mathcal{TO}}[s, o] = \sum_{t \in \mathcal{T}} C_{\mathcal{ST}}[s, t] C_{\mathcal{TO}}[t, o] \leq \max_{t \in \mathcal{T}} C_{\mathcal{TO}}[t, o].$$

In particular, this result holds for the column maximums of $C_{\mathcal{SO}}$:

$$\max_{s \in \mathcal{S}} C_{\mathcal{SO}}[s, o] \leq \max_{t \in \mathcal{T}} C_{\mathcal{TO}}[t, o].$$

Therefore, we conclude that:

$$\mathcal{ML}(C_{\mathcal{SO}}) = \log \sum_{o \in \mathcal{O}} \max_{s \in \mathcal{S}} C_{\mathcal{SO}}[s, o] \leq \log \sum_{o \in \mathcal{O}} \max_{t \in \mathcal{T}} C_{\mathcal{TO}}[t, o] = \mathcal{ML}(C_{\mathcal{TO}}).$$

□

6 Related Work

In this section, we briefly discuss some additional related work.

The problem of transmitting information through channels in cascade has been studied from the dawn of information theory, as in telecommunications it is very common to split a channel into multiple links. For the case of discrete memoryless channels with a common alphabet for the inputs and outputs, Desoer [9] proves that the Shannon capacity of a cascade of channels cannot exceed the Shannon capacity of each link in the cascade. Focusing on the same type of channels, Kiely and Coffey [14] study the effect of the *ordering* of the links on the Shannon capacity of a cascade.

The work of El-Sayed [11] provides a proof of the data processing inequality for Rényi entropies of order α (for $0 \leq \alpha \leq 1$), while we consider min-entropy, which is Rényi entropy of order ∞. Moreover, El-Sayed's definition of conditional Rényi entropy is different from the one that we use.

Alvim et al. [2] study the relationship between min-entropy leakage and *differential privacy* [10], a popular approach to protecting privacy in databases that allow statistical queries. They model a differentially-private query on a secret database S as a cascade of a deterministic channel that returns the query's real answer T (which might reveal too much about S), followed by a second channel that returns a randomized answer O. The goal is to *minimize* the leakage from S to O, $\mathcal{L}_{SO}$, while simultaneously *maximizing* the *utility* of O with respect to T, which is formalized as $V(T|O)$. We can see that our results are somehow consistent with their goals: Theorem 8 says that $\mathcal{L}_{SO} \leq \mathcal{L}_{ST}$, which means that the randomization mechanism might help but cannot hurt; and Theorem 10 says that $V(T|O) \geq V(S|O)$, which means that O's utility with respect to T may exceed but cannot be less than its utility with respect to S (which in turn corrolates closely with the leakage from S to O).

Barthe and Köpf [5] also consider the relationship between min-entropy leakage and differential privacy. Their work uses another kind of channel composition that differs from cascading—it uses *both* the original input and the intermediate result as inputs to the second channel. It is more powerful than cascading, since it drops the Markov chain restriction, but it yields a worse leakage bound. More precisely, they prove that the min-capacity of the combined channel is at most the *sum* of the min-capacities of the links, whereas with cascading it is at most the *minimum* of the min-capacities.

7 Conclusion and Future Work

In this paper, we have presented a careful account of channel cascading, and shown that cascading satisfies some nice properties with respect to min-entropy leakage. In light of the bounds on the min-entropy leakage of a cascade, we intend in future work to explore algorithms for factoring a given channel (perhaps approximately) into a cascade of channels.

Acknowledgments. This work was partially supported by the National Science Foundation under grants CNS-0831114 and CNS-1116318. We are grateful to Eduardo Ruiz for his suggestions regarding the proof of Theorem 8.

References

1. Abramson, N.: Information Theory and Coding. McGraw-Hill (1963)
2. Alvim, M., Andrés, M., Chatzikokolakis, K., Degano, P., Palamidessi, C.: Differential Privacy: On the Trade-off between Utility and Information Leakage. In: Barthe, G., Datta, A., Etalle, S. (eds.) FAST 2011. LNCS, vol. 7140, pp. 39–54. Springer, Heidelberg (2012)
3. Alvim, M., Andrés, M., Palamidessi, C.: Probabilistic information flow. In: Proc. 25th IEEE Symposium on Logic in Computer Science (LICS 2010), pp. 314–321 (2010)
4. Andrés, M., Palamidessi, C., van Rossum, P., Smith, G.: Computing the Leakage of Information-Hiding Systems. In: Esparza, J., Majumdar, R. (eds.) TACAS 2010. LNCS, vol. 6015, pp. 373–389. Springer, Heidelberg (2010)
5. Barthe, G., Köpf, B.: Information-theoretic bounds for differentially private mechanisms. In: Proc. 24th IEEE Computer Security Foundations Symposium (CSF 2011), pp. 191–204 (2011)
6. Braun, C., Chatzikokolakis, K., Palamidessi, C.: Quantitative notions of leakage for one-try attacks. In: Proc. 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009). ENTCS, vol. 249, pp. 75–91 (2009)
7. Clark, D., Hunt, S., Malacaria, P.: Quantitative information flow, relations and polymorphic types. Journal of Logic and Computation 18(2), 181–199 (2005)
8. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. John Wiley & Sons, Inc. (2006)
9. Desoer, C.A.: Communication through channels in cascade. Ph.D. thesis, Massachusetts Institute of Technology (1953)
10. Dwork, C.: A firm foundation for private data analysis. Communications of the ACM 54(1) (2011)
11. El-Sayed, A.B.: Cascaded channels and the equivocation inequality. Metrika 25, 193–208 (1978)
12. Gallager, R.G.: Information Theory and Reliable Communication. John Wiley & Sons, Inc. (1968)
13. Hamadou, S., Sassone, V., Palamidessi, C.: Reconciling belief and vulnerability in information flow. In: Proc. 31st IEEE Symposium on Security and Privacy, pp. 79–92 (2010)
14. Kiely, A.B., Coffey, J.T.: On the capacity of a cascade of channels. IEEE Transactions on Information Theory 39(4), 1310–1321 (1993)
15. Köpf, B., Basin, D.: An information-theoretic model for adaptive side-channel attacks. In: Proc. 14th ACM Conference on Computer and Communications Security (CCS 2007), pp. 286–296 (2007)
16. Köpf, B., Smith, G.: Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks. In: Proc. 23nd IEEE Computer Security Foundations Symposium (CSF 2010), pp. 44–56 (2010)
17. MacKay, D.J.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003)
18. Rényi, A.: On measures of entropy and information. In: Proc. 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960, pp. 547–561 (1961)
19. Rényi, A.: Foundations of Probability. Holden-Day, Inc. (1970)
20. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal 27, 379–423, 623–656 (1948)
21. Smith, G.: On the Foundations of Quantitative Information Flow. In: de Alfaro, L. (ed.) FOSSACS 2009. LNCS, vol. 5504, pp. 288–302. Springer, Heidelberg (2009)