Deterministic Channel Design for Minimum Leakage

Arthur Américo, MHR. Khouzani, Pasquale Malacaria
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, United Kingdom
emails: {a.passosderezende, arman.khouzani, p.malacaria}@qmul.ac.uk

***Abstract*—This work explores the problem of designing a channel that leaks the least amount of information while respecting a set of operational constraints. This paper focuses on deterministic channels and deterministic solutions. This setting is relevant because most programs and many channel design problems are naturally modelled by deterministic channels. Moreover, the setting is also relevant when considering an attacker who can observe many outputs of an arbitrary channel while the secret input stays the same: when the number of observations is arbitrarily large, the channel of minimal leakage is deterministic. The deterministic channel design problem has different solutions depending on which leakage measure is chosen. The problem is shown to be NP-hard in general. However, for a particular class of constraints, called k-complete hypergraph constraints, a greedy algorithm is shown to provide the optimal solution for a wide class of leakage measures.**

***Index Terms*—Quantitative Information Flow, Channel Design, Information Theory, Deterministic Channels**

I. Introduction

A fundamental goal of computer security is to protect secret or sensitive information from unintended disclosure. However, this goal is often unattainable, as many systems are only able to function correctly by allowing some of this information to leak to external observers. A simple password checker, for instance, always tells malicious users whether the input they provided is equal to the secret password. Even when not necessary for correct functionality of a system, it might be desirable to leak some information to increase its performance or usability. Such might be the case for systems targeted by timing attacks [23], in which an adversary is able to obtain knowledge about a secret information by observing differences in its execution time. Although it is possible to nullify this kind of attack by making the system always run for a constant time, doing so might entail an unacceptable decrease in performance for quicker tasks.

As eliminating all leakage of information is often impossible or undesirable, a natural goal is to reduce it as much as possible while guaranteeing the functionality and efficiency of the system. Despite the considerable success achieved by the field of *quantitative information flow* (QIF) in understanding and quantifying information leakage, work on how one could utilize the framework to build secure systems has been limited.

Recent work [20]–[22] shed some light on this problem, recasting it as an optimization over *information theoretic channels* – objects popularly used for abstractions of systems in QIF – bounded by some constraints, which guarantee operationality. The results obtained are promising, with a convex programming solution for a large family of design problems, and efficient solutions obtained for more specific constraints. There are, however, a myriad of other scenarios that would benefit from a similar approach.

This paper builds upon the aforementioned work by studying the design of optimal *deterministic* systems, i.e., systems whose behaviour is uniquely determined by the secret information. The motivation for exploring this problem is twofold. Firstly, this kind of problems arise in many security systems like authentication systems [28], operating systems functions [15], scheduling protocols [13], bucketing schemes in cryptography [25], anonymity (Section IV-C), and so on. A second motivation arises when the (possibly probabilistic) system to be designed may be executed a multiple number of times for a fixed secret. In this scenario, the optimal deterministic channel for a single execution is also an asymptotically optimal solution for a large number of executions.

Our optimization is taken in relation to a generalized entropy measure, introduced in [21]. This generalized measure captures most entropy measures used in the literature, such as Shannon entropy [30], min-entropy [33] and guesswork [27]. It also encompasses other proposed generalizations of information measures, such as the g-leakage framework [4] and the Rényi entropy family [29]. One desirable property in this context would be the existence of a "universal" solution, that is, a channel that minimizes the information leakage with respect to all entropy measures.

Outline and Contributions:

The main contributions of this work are: 1) to investigate the complexity of the channel design problem for deterministic channels; 2) to show that the problem has in general no universal solution; and 3) by using supermodularity, to show efficient "universal" solutions for particular class of constraints.

- Section II overviews preliminary concepts of QIF and introduces core-concave entropies.
- Section III presents the deterministic channel design problem. The problem is shown to be NP-hard, and it is proven that, in general, it does not have a universal solution.

- Section IV explores the complete k-hypergraph design problem, and proves the existence of an efficiently computable solution (a greedy algorithm), that is optimal for most entropy measures used in the literature.
- Section V considers the channel design problem for repeated executions. In this scenario, we show that the deterministic channel design problem can be used to obtain a solution that is asymptotically optimal.

A. Related Works

There is a significant body of work on generalization of Shannon entropy, e.g. Rényi entropies [29] and the more general family Sharma Mittal [31], which generalizes both Rényi and Tsallis entropies [35]. Other generalized information measures in the recent literature include the g-leakage framework [4], and a unification of the former with Rényi entropies [1]. This work uses the generalized entropy introduced in [22], which includes all aforementioned measures and also subsumes guessing entropy [27].

The problem of information leakage has been studied in quantitative information flow [8], [33] mainly in the context of measuring leakage of existing systems. Other works in the literature regarding design of security systems focus on specific contexts, such as in developing countermeasures to information leakage in timing attacks [24], [25] or in Secure Multi-Party computations [1]. Recently, a general framework for designing optimal systems has been developed [20]–[22], which considers constraints that abstract operational requirements. This paper highlights some fundamental differences from previous results in the (probabilistic) channel design problem: in previous works (only) two classes of constraints were proven to satisfy universal optimality: 1) graph constraints and 2) complete k-hypergraph constraints. In this paper, we show that in the deterministic case: 1): universal optimality fails for graph constraints (counterexample in Proposition 1). 2): for complete k-hypergraph constraints, universal optimality holds for leakage-supermodular entropies. It is true that most entropies in [20], [21] are leakage-supermodular (this is not trivial and proved in Theorem 4). Nevertheless, the latter result does not follow as an adaption of previous results: the universally optimal channel is different (greedy solution) and the proof requires a whole new approach (supermodularity).

The problem of quantifying and bounding leakage over a large number of runs has been studied in [12], [7] and [32]. Those works focus on min-entropy leakage and do not consider the problem of designing a system to minimize the leakage of information.

There is a body of work on related problems in the private information retrieval and privacy utility trade-off communities. For example [19], [34] use optimization techniques to find optimal solution to privacy leakage under constraints. However, among the many differences, they do not consider limit behaviour nor the notion of universality. Our approach also differ substantially from differential privacy [10] and ϵ-differential privacy [2], [11]. Information theoretical approaches like ours rely on statistical averages, and system executions that entail high information leakage might be tolerated if they are unlikely to occur. On the other hand, differential privacy is usually based on stronger metrics which hold in all scenarios, no matter how improbable they are.

II. Preliminaries

A. Secrets, Channels and Information Leakage

A *secret* is a piece of information the defender wishes to protect from adversaries. The secret takes values on a finite nonempty set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. The adversary may have some *knowledge* about the secret before the execution of the system. This can be modelled by a probability distribution $\pi \in \mathbb{D}\mathcal{X}$, where $\mathbb{D}\mathcal{X}$ is the set of probability distributions over $\mathcal{X}$. For brevity, π_i will be used as a shorthand for $\pi(x_i)$. Given a probability distribution $\pi \in \mathbb{D}\mathcal{X}$, the *probability vector* $\vec{\pi} \in \mathbb{R}^n$ is defined as $\vec{\pi} = (\pi_1, \ldots, \pi_n)$. The distribution π is often referred as the *prior distribution*, or simply, the *prior*.

When executed, a system takes an input from $\mathcal{X}$ and produces an *observable behaviour* (or simply *observable*) in the set $\mathcal{Y}$. Such a system can be represented by a *probabilistic discrete channel* (or simply *channel*), C which probabilistically maps inputs to outputs. The notation $C : \mathcal{X} \to \mathcal{Y}$ means that C is a channel that has $\mathcal{X}$, $\mathcal{Y}$ as input and output sets, respectively. We denote by $C(x, y)$ the conditional probability of C producing output $y \in \mathcal{Y}$ given that the input is $x \in \mathcal{X}$.

In this paper, we are mainly interested in *deterministic* channels. A channel is said to be deterministic if $\forall x \in \mathcal{X}$, $\exists! y \in \mathcal{Y}$, $C(x, y) = 1$. A deterministic channel C is completely characterized by its *characteristic function* $f_C : \mathcal{X} \to \mathcal{Y}$ defined as $f_C(x) = y$ such that $C(x, y) = 1$. Given $\mathcal{A} \subset \mathcal{X}$ and $y \in \mathcal{Y}$, the sets $f_C(\mathcal{A})$ and $f^{-1}(y)$ represent the image of $\mathcal{A}$ and the pre-image of y, respectively.

It is usually beneficial to reason about channels in a matrix representation, as in the top left of Figure 1, identifying the rows with the inputs and the columns with the outputs. For example, the channel C in Figure 1 represents a system that outputs y_2 with probability $1/3$ whenever the input is x_1. Notice that the matrix of channel C is *row-stochastic* – i.e., for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, $C(x, y) \geq 0$ and $\sum_{y' \in \mathcal{Y}} C(x, y') = 1$.

A prior π and a channel C induce a joint probability distribution $p(x, y) = \pi(x)C(x, y)$ over $\mathcal{X} \times \mathcal{Y}$. For each $y \in \mathcal{Y}$, let $p(y) = \sum_{x \in \mathcal{X}} p(x, y)$ be its marginal probability and, for each y such that $p(y) > 0$, and let $p_{\mathcal{X}|y}$ denote the probability distribution over $\mathcal{X}$ given by $p_{\mathcal{X}|y}(x) = p(x,y)/p(y)$. Let $p(x)$ and $p_{\mathcal{Y}|x}$ be similarly defined, therefore $p(x) = \pi(x)$ and $p_{\mathcal{Y}|x}(y) = C(x, y)$. The distributions $p_{\mathcal{X}|y}$ are referred to as the *posterior distributions* induced by prior π and channel C. After the execution of a system, the adversary is able to infer some information from the produced observable y, updating their knowledge about the secret from π to $p_{\mathcal{X}|y}$. An example of a channel, and the joint and posterior distributions induced by a prior is represented in Fig. 1.

In order to quantify information leakage, it is necessary to make use of an *entropy measure*, or simply *entropy*, which is a real valued function over categorical probability distributions. For each probability distribution π, the value of this function

C	y_1	y_2	y_3
x_1	$1/2$	$1/3$	$1/6$
x_2	0	$3/4$	$1/4$
x_3	0	0	1

p	y_1	y_2	y_3
x_1	$1/4$	$1/6$	$1/12$
x_2	0	$1/4$	$1/12$
x_3	0	0	$1/6$

	$p_{\mathcal{X}\mid y_1}$	$p_{\mathcal{X}\mid y_2}$	$p_{\mathcal{X}\mid y_3}$
x_1	1	$2/5$	$1/4$
x_2	0	$3/5$	$1/4$
x_3	0	0	$1/2$

Fig. 1. Top-left: a channel C with input set $\{x_1, x_2, x_3\}$ and output set $\{y_1, y_2, y_3\}$; Top-right: the corresponding joint distribution p obtained from C and the prior distribution $\vec{\pi} = (1/2, 1/3, 1/6)$; Bottom: the posterior distributions obtained from channel C and prior π. Note that channels are row stochastic matrices and posteriors are column stochastic matrices.

quantifies the *uncertainty* about the secret of an adversary whose knowledge can be described by π. By comparing the uncertainty of the adversary about the secret before and after the execution of a system, one can quantify the amount of information leakage.

The most well-known entropy measure in the literature is perhaps the *Shannon entropy* [30]. Roughly speaking, it measures the average number of yes/no set-membership questions an optimal adversary would need to ask to determine the secret value. Given $\pi \in \mathbb{D}\mathcal{X}$, its Shannon entropy is defined as:

$$H_1(\pi) = -\sum_{i=1}^{n} \pi_i \log(\pi_i),$$

with the convention that $\pi_i \log \pi_i = 0$ whenever $\pi_i = 0$. Given a prior $\pi \in \mathbb{D}\mathcal{X}$ and a channel $C : \mathcal{X} \to \mathcal{Y}$, the *posterior Shannon entropy* is defined as the expected value of Shannon entropy after the execution of the system. That is,

$$H_1(\pi, C) = \sum_{y \in \mathcal{Y}^+} p(y) H_1(p_{\mathcal{X}|y})$$

where $\mathcal{Y}^+ = \{y \in \mathcal{Y} \mid p(y) > 0\}$. The information leaked by the system execution is then taken as the *mutual information* $I_1(\pi, C) = H_1(\pi) - H_1(\pi, C)$: the reduction of the uncertainty of the adversary about the secret.

The very choice of an appropriate entropy measure depends on the specifications of the problem, and on the capabilities, actions and interests of the adversary. Two other entropy measures commonly used in the QIF literature are *guessing entropy* [27] and *min-entropy* [33]. Guessing entropy is useful to measure information in scenarios the adversary conducts an attack by brute force, as it reflects the expected number of tries an optimal attacker would need to correctly guess the secret. It is defined as $H_G(\pi) = \sum i\pi_{[i]}$, where $(\pi_{[1]}, \ldots, \pi_{[n]})$ is a non-increasing re-arrangement of $\vec{\pi}$. *Min-entropy*, given by $H_\infty(\pi) = -\log \max \pi_i$, is related to the probability that the attacker guesses the value of the secret correctly in a single try, being thus an appropriate choice for investigating one-try attack scenarios [33]. The posterior expression for guessing and min-entropy are, respectively, $H_G(\pi, C) = \sum p(y) H_G(p_{\mathcal{X}|y})$ and $H_\infty(\pi, C) = -\log \sum p(y) \max_x p_{\mathcal{X}|y}(x)$.

Given $\alpha \in \mathbb{R}_{++}$, $\alpha \neq 1$, the *Rényi entropy of order* α is defined as [29]:

$$H_\alpha(\pi) = \frac{\alpha}{1-\alpha} \log \|\vec{\pi}\|_\alpha$$

where $\|\cdot\|_\alpha$ is the α-norm function, i.e., $\|\vec{\pi}\|_\alpha = \left(\sum \pi_i^\alpha\right)^{1/\alpha}$. Shannon and min-entropy can be recovered as limit cases of the Rényi entropy family, as $\lim_{\alpha\to 1} H_\alpha(\pi) = H_1(\pi)$ and $\lim_{\alpha\to\infty} H_\alpha(\pi) = H_\infty(\pi)$. There is no general consensus on the appropriate form of posterior Rényi entropy [16]. In this paper, the posterior Rényi entropy used is the one introduced bt Arimoto [5], defined as follows

$$H_\alpha(\pi, C) = \frac{\alpha}{1-\alpha} \log \left(\sum_{y \in \mathcal{Y}} p(y) \left\|\vec{p}_{\mathcal{X}|y}\right\|_\alpha \right). \qquad (1)$$

To the extent of our knowledge, this is the only definition of posterior entropy in the literature that respects the *Data Processing Inequality* (see Section II-C) and retrieves min-entropy and Shannon entropy as limit cases.

B. A General Class of Entropies

Instead of reasoning about specific entropy measures, this work adopts a definition that generalizes most entropy measures used in the literature. This generalized entropy measure is similar to the one introduced in [22].

Definition 1: A *core-concave entropy* H is a function from categorical probability distributions to the real numbers for which there exist η, F such that:

1) For all distributions π, $H(\pi) = \eta(F(\vec{\pi}))$
2) $\eta : I \to \mathbb{R}$ is a continuous increasing real valued function defined over some interval $I \subset \mathbb{R}$.
3) F is a real valued function over probability vectors, which is concave and continuous.

Notice that for a core-concave H, the choice of η and F is not unique. In fact, given η and F for which $H(\pi) = \eta(F(\vec{\pi}))$, another choice can be obtained by taking $\eta_c(x) = \eta(cx)$ and $F_c(\vec{\pi}) = (1/c) F(\vec{\pi})$ for any $c \in \mathbb{R}_{++}$. Of course, these choices do not impact the values of $H(\pi)$, but they might affect its posterior counterpart, defined as follows.

Definition 2: Given a core-concave entropy H for a choice of η and F, its respective *posterior entropy* is

$$H(\pi, C) = \eta \left(\sum_{y \in \mathcal{Y}^+} p(y) F\left(\vec{p}_{\mathcal{X}|y}\right) \right).$$

Given $H(\pi)$ and $H(\pi, C)$, the *information leakage* (or simply *leakage*) is the difference between prior and posterior entropies, i.e.

$$\text{Leakage}(\pi, C) = H(\pi) - H(\pi, C).$$

From Definitions 1 and 2, it is possible to recover Shannon and guessing entropies by choosing $\eta(x) = x$, and respectively $F(\vec{\pi}) = -\sum \pi_i \log \pi_i$ and $F(\vec{\pi}) = \sum i\pi_{[i]}$. The Rényi entropy of order α, with its posterior form as per (1), can be recovered by choosing $\eta(x) = \frac{\alpha}{1-\alpha} \log(x)$ and $F(\vec{\pi}) = \|\vec{\pi}\|_\alpha$

when $0 < \alpha < 1$, and $\eta(x) = \frac{\alpha}{1-\alpha}\log(-x)$ and $F(\vec{\pi}) = -\|\vec{\pi}\|_\alpha$ when $\alpha > 1$. In both cases, η is increasing and F is concave.

The g-leakage framework [4] is a different approach for generalizing information leakage measures. It uses a g-vulnerability function V_g, predicated on a *gain function* $g : \mathcal{W} \times \mathcal{X} \to \mathbb{R}$, where $g(w, x)$ is the gain the adversary obtains by selecting action $w \in \mathcal{W}$ when the secret value is $x \in \mathcal{X}$. V_g is then defined simply as the expected gain of an optimal adversary, $V_g(\pi) = \max_w \sum_x \pi(x)g(w, x)$, and its posterior counterpart is naturally defined as $V_g(\pi, C) = \sum_y p(y)V_g(p_{\mathcal{X}|y})$. The generalized entropy measure given by Definition 1 subsumes the g-leakage framework: since V_g is convex for any gain function g [3], one can simply take F to be $-V_g$, and $\eta(x) = x$.

It was previously mentioned that there is no consensus on the appropriate form for conditional Rényi entropy. A different form of the posterior Rényi entropy [14] is:

$$H'_\alpha(\pi, C) = \frac{1}{1-\alpha}\log\left(\sum_{y\in\mathcal{Y}^+} p(y)\left\|\vec{p}_{\mathcal{X}|y}\right\|_\alpha^\alpha\right).$$

This alternative form can also be obtained from Definitions 1 and 2 by taking $\eta(x) = \frac{1}{1-\alpha}\log(x)$ and $F(\vec{\pi}) = \|\vec{\pi}\|_\alpha^\alpha$, for $0 < \alpha < 1$, and $\eta(x) = \frac{1}{1-\alpha}\log(-x)$ and $F(\vec{\pi}) = -\|\vec{\pi}\|_\alpha^\alpha$, for $\alpha > 1$. However, this version of posterior Rényi entropy does not converge to the posterior min-entropy as $\alpha \to \infty$. Instead, the following limit holds [17]:

$$\lim_{\alpha\to\infty} H'_\alpha(\pi, C) = -\log\left(\max_{y\in\mathcal{Y}^+,\, x\in\mathcal{X}} p_{\mathcal{X}|y}(x)\right).$$

C. Data Processing Inequality and Channel Ordering

Given channels $C : \mathcal{X} \to \mathcal{Y}$, $R : \mathcal{Y} \to \mathcal{Z}$, $D : \mathcal{X} \to \mathcal{Z}$, we write $D = CR$, and say D is the *cascading* of C and R, if:

$$D(x, z) = \sum_{y\in\mathcal{Y}} C(x, y)R(y, z), \quad \forall x \in \mathcal{X},\ z \in \mathcal{Z}.$$

In words, D is the channel obtained by feeding the observable produced by C as input to R. It is also said that the channel D is obtained by *post-processing* the output of C by R. Notice that, if we consider matrix representations, the matrix of D is obtained by multiplying the matrix of C by R.

Cascading induces a pre-order $\sqsupseteq_\circ$ over channels with a same input set, in which $C \sqsupseteq_\circ D$ iff there is R such that $D = CR$. If $C \sqsupseteq_\circ D$ and $D \sqsupseteq_\circ C$, we write $C =_\circ D$.

Notice that, if $D = CR$, an adversary that has access to the behaviour of C can trivially recover D simply by feeding the observable generated by C to R. Thus, intuitively channel D would leak at most as much information as C. This intuition is formalised in the following result, which can be seen as a generalization of the "data processing inequality" for core-concave entropies.

Theorem 1: Let $C : \mathcal{X} \to \mathcal{Y}$, $D : \mathcal{X} \to \mathcal{Z}$. If $C \sqsupseteq_\circ D$, then for any $\pi \in \mathbb{D}\mathcal{X}$ and any core-concave entropy H, $H(\pi, C) \leq H(\pi, D)$.

Proof: Since $p(y) = \sum_{z\in\mathcal{Z}^+} p(z)p(y|z)$, we can write:

$$\sum_{y\in\mathcal{Y}^+} p(y)F\left(\vec{p}_{\mathcal{X}|y}\right) = \sum_{y\in\mathcal{Y}^+}\left(\sum_{z\in\mathcal{Z}^+} p(z)p(y|z)\right)F\left(\vec{p}_{\mathcal{X}|y}\right)$$
$$= \sum_{y\in\mathcal{Y}^+, z\in\mathcal{Z}^+} p(z)p(y|z)F\left(\vec{p}_{\mathcal{X}|y,z}\right)$$
$$\leq \sum_{z\in\mathcal{Z}^+} p(z)F\left(\sum_{y\in\mathcal{Y}^+} p(y|z)\vec{p}_{\mathcal{X}|y,z}\right) = \sum_{z\in\mathcal{Z}^+} p(z)F\left(\vec{p}_{\mathcal{X}|z}\right)$$

The second equality follows because $D = CR$ implies $\vec{p}_{\mathcal{X}|y} = \vec{p}_{\mathcal{X}|y,z}$. Next, Jensen's inequality is applied for concave F, noting that $p(y|z)$ for $y \in \mathcal{Y}^+$ constitute convex coefficients. The last equality uses $\sum_{y\in\mathcal{Y}^+} p(y|z)\vec{p}_{\mathcal{X}|y,z} = \vec{p}_{\mathcal{X}|z}$. The lemma follows by applying $\eta(\cdot)$ to the steps, noting that η preserve the inequality since it is increasing. ■

An important consequence of the corollary below is that leakage is always non-negative:

Corollary 1: For any core-concave entropy H, prior π and channel C, $H(\pi) \geq H(\pi, C)$.

Proof: Consider the channel $\overline{0} : \mathcal{X} \to \mathcal{Y}_1$ where $\mathcal{Y}_1 = \{y\}$ is a singleton and $\overline{0}(x, y) = 1$ for all $x \in \mathcal{X}$. It follows from Definition 2 that $H(\pi) = H(\pi, \overline{0})$. Given any channel $C : \mathcal{X} \to \mathcal{Z}$, define $R_C : \mathcal{Z} \to \mathcal{Y}_1$ as $R_C(z, y) = 1$ for all $z \in \mathcal{Z}$. Then, $\overline{0} = CR_C$, and the claim follows from Theorem 1. ■

III. The Deterministic Channel Design Problem

Given an input set $\mathcal{X}$ and an output set $\mathcal{Y}$, the deterministic channel design problem is to find a deterministic channel $C : \mathcal{X} \to \mathcal{Y}$ that minimizes leakage with respect to a prior π and a core-concave entropy H, while respecting some operational constraints. As the information leakage is $H(\pi) - H(\pi, C)$ and π is assumed given, the problem simplifies to finding such C that maximizes $H(\pi, C)$. Moreover, since η is increasing, the problem further simplifies into maximizing $\sum_{y\in\mathcal{Y}^+} p(y)F(\vec{p}_{\mathcal{X}|y})$.

The operational constraints are of two types. The first type, called *hard* constraints, are defined by a set $\Omega \subset \mathcal{X} \times \mathcal{Y}$. They define the strict behaviour of the system: (x, y) is an element of Ω if and only if the observable y can be generated when the secret value is x. The set of channels that respect Ω can thus be defined as:

$$\Gamma_\Omega = \{C : \mathcal{X} \to \mathcal{Y} \mid C(x, y) > 0 \implies (x, y) \in \Omega\}.$$

The second type of constraint, referred to as the *soft* constraint, is predicated on a *utility function* $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, which represents how useful or desirable each behaviour of the system is given each secret value. The expected value of u, given by $\mathbb{E}_{\pi,C}[u] = \sum_{x,y} \pi(x)C(x, y)u(x, y)$, reflects the average performance of the system, such as execution time. The desirable outcome is to obtain a channel C that minimizes the leakage of information, while ensuring an acceptable performance. This is accomplished by adding $\mathbb{E}[u] \geq u_{min}$ as a constraint of our design problem, for an appropriate threshold $u_{min} \in \mathbb{R}$.

Definition 3: The *deterministic channel design problem* is to find $C \in \Gamma_\Omega$ that maximizes

$$\sum_{y \in \mathcal{Y}^+} p(y) F(\vec{p}_{\mathcal{X}|y})$$

subject to C being deterministic and $\mathbb{E}_{\pi,C}[u] \geq u_{min}$.

Notice that our definition allows for constraints that do not limit the space of possible solutions. Namely, one could define $\Omega = \mathcal{X} \times \mathcal{Y}$, or u to be identically 1 and $u_{min} = 0$. Therefore, Definition 9 allows us to define channel design problems predicated only on either hard or soft constraints.

A more general version of this problem was studied in [20], [22], in which the requirement of channels being deterministic is lifted. It is shown in that the (probabilistic) channel design problem can be solved by convex optimization techniques. In particular, the KKT conditions are necessary and sufficient for a solution, for any choice of core-concave entropy.

A. Complexity results

The first result on this section regards the computational complexity of the deterministic channel design problem. In particular, we show this is NP-hard. This is achieved by proving that a related decision problem is NP-complete. It suffices to consider min-entropy as the entropy measure, and only hard constraints.

Theorem 2: Given an input set $\mathcal{X}$, an output set $\mathcal{Y}$, a prior $\pi \in \mathbb{D}\mathcal{X}$, a set of hard constraints $\Omega \subset \mathcal{X} \times \mathcal{Y}$ and a threshold $t \in \mathbb{R}_+$, it is NP-complete to decide whether there is a deterministic channel $C \in \Gamma_\Omega$ such that $H_\infty(\pi, C) \geq t$.

Proof: First notice that, given C, we can calculate $H_\infty(\pi, C) = -\log \sum_y \max_x \pi(x) C(x, y)$ in $O(|\mathcal{X}||\mathcal{Y}|)$ time. Thus, the problem is in NP.

NP-completeness is proven by reduction from the set cover problem (SCP) [18], which is defined as follows: Given a set $\mathcal{U}$, a collection $\mathcal{C} \subset 2^{\mathcal{U}}$ such that $\bigcup_{S \in \mathcal{C}} S = \mathcal{U}$ and an integer k, decide whether there exists a sub-collection $\mathcal{C}'$ of sets of $\mathcal{C}$ such that $\bigcup_{S \in \mathcal{C}'} S = \mathcal{U}$ and $|\mathcal{C}'| \leq k$.

Given an instance $(\mathcal{U}, \mathcal{C}, k)$ of the SCP, we reduce it to an instance of the decision problem stated in the theorem by the following transformation

$$\mathcal{X} = \mathcal{U}, \quad \mathcal{Y} = \mathcal{C}, \quad \pi(x) = \frac{1}{|\mathcal{U}|} \text{ for all } x \in \mathcal{U}$$

$$\Omega = \{(x, y) \in \mathcal{U} \times \mathcal{C} \mid x \in y\}, \quad t = -\log \frac{k}{|\mathcal{U}|}.$$

We prove that $(\mathcal{U}, \mathcal{C}, k)$ satisfies the SCP if and only if $(\mathcal{X}, \mathcal{Y}, \pi, \Omega, t)$ defined above satisfies the decision problem described in the theorem. Suppose $(\mathcal{U}, \mathcal{C}, k)$ satisfies the set cover problem. Enumerate the elements of $\mathcal{C}$ as $\mathcal{C} = \{S_1, S_2, \ldots S_{|\mathcal{C}|}\}$. Let $\mathcal{C}' \subset \mathcal{C}$ be a sub-collection of size at most k that covers $\mathcal{U}$. Define $C : \mathcal{U} \to \mathcal{C}$ as follows.

$$C(u, S_i) = \begin{cases} 1, & \text{if } u \in S_i, S_i \in \mathcal{C}' \\ & \text{and } \forall j < i, S_j \in \mathcal{C}' \Rightarrow u \notin S_j, \\ 0, & \text{otherwise.} \end{cases}$$

By construction, $C(u, S_i) > 0 \Rightarrow u \in S_i \Rightarrow (u, S_i) \in \Omega$, so $C \in \Gamma_\Omega$. Since π is uniform,

$$H_\infty(\pi, C) = -\log \frac{1}{|\mathcal{U}|} \sum_{S \in \mathcal{C}} \max_{u \in \mathcal{U}} C(u, S) \geq -\log \frac{|\mathcal{C}'|}{|\mathcal{U}|}.$$

As $|\mathcal{C}'| \leq k$, we obtain $H_\infty(\pi, C) \geq t$.

Conversely, suppose that there is a deterministic channel $C \in \Gamma_\Omega$ such that $H_\infty(\pi, C) \geq t$. Let $\mathcal{C}' = \{S \in \mathcal{C} \mid \exists u \in \mathcal{U};\ C(u, S) = 1\}$. Then,

$$t = -\log \frac{k}{|\mathcal{U}|} \geq H_\infty(\pi, C) \geq -\log \frac{|\mathcal{C}'|}{|\mathcal{U}|},$$

and we obtain $|\mathcal{C}'| \leq k$. Also, $\forall u \in \mathcal{U}$, $\exists S \in \mathcal{C}'$ such that $C(u, S) = 1$, and therefore $u \in \bigcup_{S \in \mathcal{C}'} S$. Thus, $(\mathcal{U}, \mathcal{C}, k)$ satisfies the SCP. ■

B. Universality of the solution

As mentioned earlier, the choice of the appropriate entropy measure is not trivial, and it depends on the interests of the adversary, as well as on his capabilities and attack strategies. As these factors are often outside the control of the designer of the system, one desirable attribute of channel design problems would be if they are *universal* with regards to the choice of entropy measure – i.e., having all other variables fixed, they admit a channel that is a solution to the problem for any choice of the entropy.

The probabilistic version of the channel design problem does not, in general, satisfy universality [20]. In fact, and perhaps unsurprisingly, universality is rarely achievable. Currently, the existence of universal solutions for the probabilistic case have been proven for only two classes of hard constraints: "cloaking constraints" (studied in next section as complete k-hypergraph constraints) [21] and "graph constraints" [20] – in which at most two observables can be produced by each secret.

Next result shows that, in general, universality is also not satisfied in the deterministic channel design problem. The proof also demonstrates that in the deterministic case, contrary to the probabilistic one [20], [21], there is no universal solution even for "graph constraints".

Proposition 1: In general, there is no channel that is a solution for a deterministic channel design problem for all entropy measures.

Proof: Consider the deterministic channel design problem given by $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, $\mathcal{Y} = \{\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_4\}\}$, $\Omega = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid x \in y\}$ and $\vec{\pi} = (0.35, 0.35, 0.15, 0.15)$. One can verify that Shannon and Min-entropy have different optimal solutions, respectively, C_{Sh} and C_{Me} represented below.

C_{Sh}	$\{x_1, x_3\}$	$\{x_2, x_4\}$
x_1	1	0
x_2	0	1
x_3	1	0
x_4	0	1

C_{Me}	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_4\}$
x_1	1	0	0
x_2	1	0	0
x_3	0	1	0
x_4	0	0	1

In particular, while for Shannon entropy $H_1(\pi; C_{Me}) = 0.7$ and $H_1(\pi; C_{Sh}) \approx 0.88$, for min-entropy $H_\infty(\pi; C_{Me}) \approx 0.62$ and $H_\infty(\pi; C_{Sh}) \approx 0.51$. ■

Notice that for graph constraints and a uniformly distributed prior, the deterministic channel design problem for min entropy is equivalent to the *edge cover* problem, for which there exists a polynomial solution. For example, consider $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $\mathcal{Y} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_1, x_5\}, \{x_2, x_5\}, \{x_3, x_6\}, \{x_5, x_6\}\}$, $\Omega = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid x \in y\}$ and $\vec{\pi} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$; then the optimal solution is $\{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}\}$.

IV. Universality for Deterministic Channels: The complete k-hypergraph design problem

Last section outlined two undesirable characteristics of the deterministic channel design problem. The first one is its NP-hardness, which makes it computationally difficult to find a solution. The second one is that a universal solution – i.e., one that is optimal for all entropy measures – does not exist in general. One natural course of action is to explore subsets of the deterministic channel design problem that admit either an efficient or a universal solution. Both these characteristics can be found in the *complete k-hypergraph design problem*, as we investigate in this section.

Definition 4: The *complete k-hypergraph design problem* is a deterministic channel design problem as in Definition 3 in which, $\mathcal{Y} = \{\mathcal{A} \subset \mathcal{X} \mid |\mathcal{A}| = k\}$ and $\Omega = \{(x, y) \mid x \in y\}$.

At a high level, the problem models a scenario in which each observable can be produced only by at most k secrets. A variant of this problem, for the probabilistic case, has been studied in [21] under the name of cloaking constraints. They have proven that for all k, there exists a solution that is optimal for any choice of symmetric entropies.

Here, we prove a similar result for the deterministic case. For that end, however, it is necessary to restrict the entropy measures under consideration. Besides being core-concave, we also require that they be *symmetric* and *expansible*.

Definition 5: A core concave entropy H is said to be

1) *Symmetric*: if, for all $\vec{\pi} = (\pi_1, \dots \pi_n)$ and all permutations ϕ over $\{1, \dots, n\}$, $F(\vec{\pi}) = F((\pi_{\phi(1)}, \dots, \pi_{\phi(n)}))$
2) *Expansible*: given $\vec{\pi} = (\pi_1, \dots, \pi_n)$ and $\vec{\pi}' = (\pi_1, \dots, \pi_n, 0)$, we have $F(\vec{\pi}) = F(\vec{\pi}')$.

Most of the entropies used in the literature are symmetric and expansible, including all entropy measures defined in Section II-A.

For the remainder of this section, let $k \in \mathbb{N}_{++}$, $\mathcal{X} = \{x_1, \dots, x_n\}$, $\pi \in \mathbb{D}\mathcal{X}$, $\mathcal{Y} = \{\mathcal{A} \subset \mathcal{X} \mid |\mathcal{A}| = k\}$ and $\Omega = \{(x, y) \mid x \in y\}$. Without loss of generality, assume the elements of $\mathcal{X}$ are labelled in a non-increasing order, i.e., we have $\pi_i \geq \pi_j$ whenever $i < j$.

Algorithm 1 proposes a greedy solution $C_k \in \Gamma_\Omega$ for the complete k-hypergraph design problem, returning its characteristic function f_{C_k}. The algorithm is very simple: it takes the k most likely secrets and associate them with the first observable, then takes the k most likely secrets among the remaining secrets and associate them with the second observable, and so on. If n is not divisible by k, the last $(n \mod k)$ input values should be mapped to any permissible output. In line 5, Algorithm 1 selects one such output by completing the set with the first output values until it reaches size k.

Algorithm 1 Greedy algorithm for the k-complete hypergraph problem

```
    Input: Input set $\mathcal{X}$, prior $\pi \in \mathbb{D}\mathcal{X}$ and integer $k \leq |\mathcal{X}|$
    Output: A characteristic function $f_{C_k} : \mathcal{X} \to \mathcal{Y}$ .
1:  for $i \in \{0, \dots, \lfloor n/k \rfloor - 1\}$ do
2:      $y_{i+1} \leftarrow \{x_{ik+1}, x_{ik+2}, \dots, x_{ik+(k-1)}, x_{(i+1)k}\}$,
3:  $r \leftarrow \lceil n/k \rceil$
4:  if $n/k < r$ then
5:      $y_r \leftarrow \{x_{rk+1}, \dots, x_n, x_1, x_2, \dots, x_{rk-n}\}$
6:  for $j \in \{1, \dots, n\}$ do
7:      $f_{C_k}(x_j) \leftarrow y_{\lceil j/n \rceil}$
8:  return $f_{C_k}$
```

C_k	y_1	y_2	y_3
x_1	1	0	0
x_2	1	0	0
x_3	1	0	0
x_4	0	1	0
x_5	0	1	0
x_6	0	1	0
x_7	0	0	1

Fig. 2. The solution given by Algorithm 1 for $k = 3$.

A. Entropies and Supermodular Functions

This section introduces the concept of *leakage-supermodular* entropies. The main result, Theorem 5, proves that Algorithm 1 is optimal for all such entropies.

We begin with supermodularity, a concept also known in the literature as L-superadditivity. For a complete treatment of supermodularity, one can refer to Chapter 6 of [26].

Let $\leq_c$ be the component-wise partial order over $\mathbb{R}^n_+$, defined as follows: given two vectors $\mathbf{p} = (p_1, \dots, p_n)$ and $\mathbf{q} = (q_1 \dots, q_n)$ in $\mathbb{R}_+$, $\mathbf{p} \leq_c \mathbf{q}$ iff $p_i \leq q_i$ for all i. The set $\mathbb{R}^n_+$ and the partial order $\leq_c$ define a lattice in which the *join* and *meet* of two elements are given by

$$\mathbf{p} \vee \mathbf{q} = (\max(p_1, q_1), \dots, \max(p_n, q_n)), \qquad \text{and}$$
$$\mathbf{p} \wedge \mathbf{q} = (\min(p_1, q_1), \dots, \min(p_n, q_n))$$

Definition 6: A function $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ is *supermodular* if for all $\mathbf{p} = (p_1, \dots, p_n)$ and $\mathbf{q} = (q_1, \dots, q_n)$

$$\phi(\mathbf{p} \vee \mathbf{q}) + \phi(\mathbf{p} \wedge \mathbf{q}) \geq \phi(\mathbf{p}) + \phi(\mathbf{q}).$$

An equivalent definition of supermodular functions relies in comparing the entries of the function two by two [26]. Let $\{\mathbf{e_i}\}_{i\in\{0,\dots,n\}}$ be the canonical basis of $\mathbb{R}^n$. Then:

Definition 7: A function $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ is *supermodular* if for all $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}^n_+$, all $i, j \leq n$ such that $i \neq j$, and all $\delta_1, \delta_2 \geq 0$, we have

$$\phi(\mathbf{p} + \delta_1\mathbf{e_i} + \delta_2\mathbf{e_j}) + \phi(\mathbf{p}) \geq \phi(\mathbf{p} + \delta_1\mathbf{e_i}) + \phi(\mathbf{p} + \delta_2\mathbf{e_j}). \quad (2)$$

The following Theorem, known as Topkis's characterization theorem, will be useful for characterizing supermodular functions in some of our next results [26].

Theorem 3: Suppose $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ have second partial derivatives. Then ϕ is supermodular if and only if for all $i, j \leq n$ and $i \neq j$,

$$\frac{\partial^2\phi(\mathbf{p})}{\partial p_i \partial p_j} \geq 0$$

We are now in position to derive the results of this section, which rely on the above characterizations of supermodular functions. These results, however, are applicable only to a subset of core-concave entropies, which we call *leakage-supermodular*.

Definition 8: Let H be an expansible and symmetric core-concave entropy for a given η and F. Define G_F as follows: for all $n \geq 1$ and all $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}^n_+$,

$$G_F(\mathbf{p}) = \left(\sum_i^n p_i\right) F\left(\frac{p_1}{\sum_i^n p_i}, \dots, \frac{p_n}{\sum_i^n p_i}\right).$$

and $G_F(\mathbf{p}) = 0$ whenever all entries of $\mathbf{p}$ are 0. Then we say H is *leakage-supermodular* if for all $n \geq 2$, the restriction of G_F to $\mathbb{R}^n_+$ is supermodular.

The functions G_F defined above are closely related to the value of the posterior entropy. In fact, one can calculate $H(\pi, C)$ by applying G_F to the joint probability distribution $p(x, y) = \pi(x)C(x, y)$.

Proposition 2: For any expansible and symmetric core-concave entropy H, prior π and channel C,
$H(\pi, C) = \eta\left(\sum_{y\in\mathcal{Y}} G_F\left(p(x_1, y), \dots, p(x_n, y)\right)\right)$

Proof: Immediate from Definitions 2 and 8. ■

Example 1: Consider the following joint probability distribution, derived by the channel C in Figure 1 and prior $(1/2, 1/3, 1/6)$:

p	y_1	y_2	y_3
x_1	$1/4$	$1/6$	$1/12$
x_2	0	$1/4$	$1/12$
x_3	0	0	$1/6$

Proposition 2 says that $H(\pi, C)$ is equal to

$$\eta(G_F(1/4, 0, 0) + G_F(1/6, 1/4, 0) + G_F(1/12, 1/12, 1/6)).$$

The equivalence in Proposition 2 gives an important connection between supermodular functions and posterior forms of leakage-supermodular entropies, allowing us to apply the extensive literature regarding the supermodularity to our framework. The following result establishes a fundamental connection between supermodularity and entropy:

Theorem 4: Rényi-entropies (including Shannon and min-entropy) and guessing entropy are leakage-supermodular.

Proof: As they are defined in the limits for $\alpha \to 1$ or $\alpha \to \infty$, we provide a separate proof for Shannon and min-entropy.

Throughout this proof, let $\pi \in \mathbb{D}\mathcal{X}$, $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_+$ and $S_\mathbf{p} = \sum_i^n p_i$. Also let $i, j \leq n$ with $i \neq j$ and $\delta_1, \delta_2 \geq 0$.

Min-Entropy: For min-entropy, we have $G_F(\mathbf{p}) = -S_\mathbf{p} \max_i p_i/S_\mathbf{p} = -\max_i p_i$. Clearly, it must be that $G_F(\mathbf{p} \vee \mathbf{q}) = G_F(\mathbf{p})$ or $G_F(\mathbf{p} \vee \mathbf{q}) = G_F(\mathbf{q})$. Suppose, without loss of generality, that $G_F(\mathbf{p} \vee \mathbf{q}) = G_F(\mathbf{p})$. As $G_F(\mathbf{q}) \leq G_F(\mathbf{p} \wedge \mathbf{q})$, we obtain

$$G_F(\mathbf{p} \vee \mathbf{q}) + G_F(\mathbf{p} \wedge \mathbf{q}) \geq G_F(\mathbf{p}) + G_F(\mathbf{q})$$

and G_F is supermodular.

Guessing-Entropy: For Guessing-Entropy, we have $G_F(\mathbf{p}) = S_\mathbf{p} \sum_i p_{[i]}/S_\mathbf{p} = \sum_i p_{[i]}$, where $(p_{[1]}, \dots, p_{[n]})$ is a non-increasing rearrangement of $\mathbf{p}$. Suppose, without loss of generality, that $p_i + \delta_1 \geq p_j + \delta_2$. We will prove that

$$G_F(\mathbf{p} + \delta_1\mathbf{e_i} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p} + \delta_1\mathbf{e_i}) \geq G_F(\mathbf{p} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p}). \quad (3)$$

Let $\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}} = \{k \leq n \mid p_j < p_k < p_j + \delta_2\}$ – that is, $\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$ is the set of coordinates of $\mathbf{p} + \delta_1\mathbf{e_i}$ whose value is between p_j and $p_j + \delta_2$. Suppose $p_j + \delta_2$ is the mth biggest element of the vector $\mathbf{p} + \delta_1\mathbf{e_i} + \delta_1\mathbf{e_j}$. Then, p_j is the $(m + |\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|)$th biggest element of $\mathbf{p} + \delta_1\mathbf{e_i}$.

The difference $G_F(\mathbf{p} + \delta_1\mathbf{e_i} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p} + \delta_1\mathbf{e_i})$ is then the difference of the contribution of the jth coordinate, plus the sum of the entries given by the set $\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$, as the integers by which they are multiplied are decreased by one. As we assume $p_i + \delta_1 \geq p_j + \delta_2$, $i \notin \mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$.

$$\begin{aligned} &G_F(\mathbf{p} + \delta_1\mathbf{e_i} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p} + \delta_1\mathbf{e_i}) \\ =&m(p_j + \delta_2) + \sum_{k\in\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}} p_k - (m + |\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|)p_j \\ =&-|\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|p_j + m\delta_2 + \sum_{k\in\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}} p_k. \end{aligned}$$

We now turn to the right hand side of (3), the difference $G_F(\mathbf{p} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p})$. We define $\mathcal{I}_\mathbf{p}$ similarly to $\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$ and divide the proof in three cases.

Case 1 ($p_i \geq p_j + \delta_2$): In this case, $p_j + \delta_2$ is the mth biggest element of $\mathbf{p} + \delta_2\mathbf{e_j}$ and p_j the $(m + |\mathcal{I}_\mathbf{p}|)$th biggest element of $\mathbf{p}$. Moreover, $\mathcal{I}_\mathbf{p} = \mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$, and we obtain

$$\begin{aligned} &G_F(\mathbf{p} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p}) \\ =&m(p_j + \delta_2) + \sum_{k\in\mathcal{I}_\mathbf{p}} p_k - (m + |\mathcal{I}_\mathbf{p}|)p_j \\ =&G_F(\mathbf{p} + \delta_1\mathbf{e_i} + \delta_2\mathbf{e_j}) - G_F(\mathbf{p} + \delta_1\mathbf{e_i}). \end{aligned}$$

Case 2 ($p_j < p_i < p_j + \delta_2$): In this case, $p_j + \delta_2$ is the $(m - 1)$th biggest element of $\mathbf{p} + \delta_2\mathbf{e_j}$, and p_j is the $(m - 1 + |\mathcal{I}_\mathbf{p}|)$th biggest element of $\mathbf{p}$. This time, $\mathcal{I}_{\mathbf{p}} = \mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}} \cup \{i\}$, and we have

$$\begin{aligned}
&G_F(\mathbf{p}+\delta_2\mathbf{e_j}) - G_F(\mathbf{p})\\
=&(m-1)(p_j+\delta_2) + \sum_{k\in\mathcal{I}_{\mathbf{p}}} p_k - (m-1+|\mathcal{I}_{\mathbf{p}}|)p_j\\
=&(m-1)(p_j+\delta_2) + \sum_{k\in\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}} p_k + p_i - (m+|\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|)p_j\\
=&-|\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|p_j + m\delta_2 + \sum_{k\in\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}} p_k + p_i - (p_j+\delta_2)\\
\leq&-|\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}|p_j + m\delta_2 + \sum_{k\in\mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}} p_k\\
=&G_F(\mathbf{p}+\delta_1\mathbf{e_i}+\delta_2\mathbf{e_j}) - G_F(\mathbf{p}+\delta_1\mathbf{e_i}),
\end{aligned}$$

where the last inequality comes from the assumption that $p_i < p_j + \delta_2$.

Case 3 ($p_i \leq p_j$): In this case, we again have that $p_j + \delta_2$ is the $(m-1)$th biggest element of $\mathbf{p}+\delta_2\mathbf{e_j}$, and p_j is the $(m-1+|\mathcal{I}_{\mathbf{p}}|)$th biggest element of $\mathbf{p}$. This time, $\mathcal{I}_{\mathbf{p}} = \mathcal{I}_{\mathbf{p}+\delta_1\mathbf{e_i}}$, and we have

$$\begin{aligned}
&G_F(\mathbf{p}+\delta_2\mathbf{e_j}) - G_F(\mathbf{p})\\
=&(m-1)(p_j+\delta_2) + \sum_{k\in\mathcal{I}_{\mathbf{p}}} p_k - (m-1+|\mathcal{I}_{\mathbf{p}}|)p_j\\
=&-|\mathcal{I}_{\mathbf{p}}|p_j + (m-1)\delta_2 + \sum_{k\in\mathcal{I}_{\mathbf{p}}} p_k\\
\leq&-|\mathcal{I}_{\mathbf{p}}|p_j + m\delta_2 + \sum_{k\in\mathcal{I}_{\mathbf{p}}} p_k\\
=&G_F(\mathbf{p}+\delta_1\mathbf{e_i}+\delta_2\mathbf{e_j}) - G_F(\mathbf{p}+\delta_1\mathbf{e_i}).
\end{aligned}$$

Therefore, G_F is supermodular.

Shannon Entropy: For Shannon Entropy, $G_F(\mathbf{p}) = S_{\mathbf{p}}\sum_i^n {}^{p_i}\!/\!{}_{S_{\mathbf{p}}} \log\left(\frac{S_{\mathbf{p}}}{p_i}\right) = \sum_i^n p_i \log\left(\frac{S_{\mathbf{p}}}{p_i}\right)$. Consider an i, j, $i \neq j$. If $p_i, p_j > 0$, we have

$$\frac{\partial^2 G_F(\mathbf{p})}{\partial p_i \partial p_j} = \frac{1}{S_{\mathbf{p}} \ln(2)} \geq 0. \tag{4}$$

Supermodularity then follows from Theorem 3. If, on the other hand, p_i or p_j are equal to 0, the second derivative does not exist. For such a case, let $\epsilon > 0$ and $\mathbf{p}' = \mathbf{p} + \epsilon(\mathbf{e_i} + \mathbf{e_j})$. Then Theorem 3 and (4) imply that

$$\begin{aligned}
G_F(\mathbf{p}'+\delta_1\mathbf{e_i}+\delta_2\mathbf{e_j}) + G_F(\mathbf{p}') \geq&\\
G_F(\mathbf{p}'+\delta_1\mathbf{e_i}) &+ G_F(\mathbf{p}'+\delta_2\mathbf{e_j}).
\end{aligned}$$

Taking $\epsilon \to 0$, we can conclude that G_F is supermodular following Definition 7.

Rényi Entropies: For Rényi entropy of order α, we have $G_F(\mathbf{p}) = S_{\mathbf{p}}\|({}^1\!/\!{}_{S_{\mathbf{p}}})\mathbf{p}\|_a = \|\mathbf{p}\|_a$ if $\alpha < 1$ and $G_F(\mathbf{p}) = -\|\mathbf{p}\|_\alpha$ if $\alpha > 1$. For an i, j, $i \neq j$ where $p_i, p_j > 0$, we have

$$\frac{\partial^2 \|\mathbf{p}\|}{\partial p_i \partial p_j} = (1-\alpha)\, p_i^{\alpha-1} p_j^{\alpha-1} \left(\sum_{k=1}^n p_k^\alpha\right)^{\frac{1}{\alpha}-2}$$

which is negative for $\alpha > 1$ and positive for $\alpha < 1$, and supermodularity again follows from Theorem 3. When p_i or p_j is 0 where the derivatives may not exist, supermodularity follows from a similar argument as for the Shannon Entropy. ■

Leakage-supermodularity, however, is not shared by all symmetric, expansible core concave entropies.

Proposition 3: Not all core-concave entropies are leakage-supermodular

Proof: In Proposition 4, we considered the choice of η and F that yield the posterior Rényi entropy of the form (1). As discussed in Section II-B, another form of posterior entropy can be obtained by taking (for $\alpha > 1$) $\eta(x) = ({}^1\!/\!{}_{1-\alpha})\log(-x)$ and $F(\vec{\pi}) = -\|\vec{\pi}\|_\alpha^\alpha$.

This alternative form, however, is not leakage-supermodular. Consider $\mathbf{p} = (0.15, 0.15, 0.5)$, and $\mathbf{q} = (0.25, 0.05, 0.5)$. Then, for $\alpha = 2$, we obtain

$$\begin{aligned}
G_F(\mathbf{p} \vee \mathbf{q}) + G_F(\mathbf{p} \wedge \mathbf{q}) &\approx -0.765, \quad \text{and}\\
G_F(\mathbf{p}) + G_F(\mathbf{q}) &\approx -0.763.
\end{aligned}$$

■

B. Universality for Leakage-supermodular Entropies

In the discussion above, it was claimed that the greedy solution given by Algorithm 1 is optimal for all leakage-supermodular entropies. This section is devoted to proving this claim.

Theorem 5: Given a complete k-hypergraph channel design problem the solution given by Algorithm 1 is optimal for any leakage-supermodular entropy.

Proof: The proof presented here takes a more intuitive approach, helpful for visualizing the application of the concepts of this section (a more formal version of the proof is provided in Appendix A). For the proof of the theorem, it will be helpful to reason in terms of the joint probability matrix, as in Example 1.

Let C_k be the channel obtained by Algorithm 1. The joint matrix J_k has the following form:

$$\begin{pmatrix}
\pi_1 & 0 & \dots & 0\\
\vdots & \vdots & & \vdots\\
\pi_k & 0 & & \vdots\\
0 & \pi_{k+1} & & \vdots\\
\vdots & \vdots & & \vdots\\
\vdots & \pi_{2k} & & \vdots\\
\vdots & 0 & & \vdots\\
\vdots & \vdots & \ddots & \vdots\\
\vdots & \vdots & & 0\\
\vdots & \vdots & & \pi_{mk+1}\\
\vdots & \vdots & & \vdots\\
0 & 0 & \dots & \pi_n
\end{pmatrix}$$

Suppose J is another joint matrix that respects the complete k-hypergraph problem constraints for the same prior. We now describe an algorithm to derive J_k from J such that each step increases entropy.

A step in the algorithm uses the following three sub-steps:

- select and align two columns c_i, c_j
- perform $\wedge, \vee$ operations on the aligned columns and replace c_i, c_j with $c_i \vee c_j, c_i \wedge c_j$
- dis-align the two columns $c_i \vee c_j, c_i \wedge c_j$

Let's illustrate one step with the following example:

Consider the following joint matrix, and suppose $k = 2$:

$$\begin{pmatrix} 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \\ 0 & 0.15 & 0 \\ 0 & 0 & 0.1 \\ 0.05 & 0 & 0 \end{pmatrix}$$

We select the two columns which contain the two most likely priors (i.e. 0.4,0.3); these are columns 2 and 3; we align c_2, c_3 so that 0.4,0.3 appear in different rows resulting in:

$$\begin{pmatrix} 0 & 0.4 & 0.1 \\ 0 & 0.15 & 0.3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.05 & 0 & 0 \end{pmatrix}$$

we then replace c_2, c_3 with $c_2 \vee c_3, c_2 \wedge c_3$ obtaining

$$\begin{pmatrix} 0 & 0.4 & 0.1 \\ 0 & 0.3 & 0.15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.05 & 0 & 0 \end{pmatrix}$$

finally we dis-align the columns 2 and 3, i.e. position values in $c_2 \vee c_3, c_2 \wedge c_3$ so that each row has the same probability it had before the step, obtaining:

$$\begin{pmatrix} 0 & 0.4 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.15 \\ 0 & 0 & 0.1 \\ 0.05 & 0 & 0 \end{pmatrix}$$

Notice that the posterior entropy of the joint matrix J is $\eta(\sum_i G_F(c_i))$ where the c_i's are the columns of the matrix J and since G_F is symmetric, aligning columns doesn't change the posterior, i.e. $G_F(c_i) = G_F(c'_i)$ for any permutation c'_i of the column c_i. Same holds for dis-aligning columns, because again it is a permutation. Let's consider the remaining sub-step, i.e. replacing c_i, c_j with $c_i \vee c_j, c_i \wedge c_j$. By supermodularity of G we have $G_F(c_i) + G_F(c_j) \leq G_F(c_i \vee c_j) + G_F(c_i \wedge c_j)$ hence that sub-step increases (or keeps equal) the posterior entropy. Notice that at the conclusion of the step, the new matrix has the same probabilities in each row it had before that step, hence it is still a joint matrix that respects the complete k-hypergraph problem constraints for the same prior.

The selection and alignment of columns is as follows: at the initial step select c_i such that c_i contains the first r elements with the highest probabilities, say $\pi_1, \ldots, \pi_r$; if $r < k$ then select c_j as the column containing π_{r+1}; align c_i, c_j so that π_{r+1} is not on the same row as any of the $\pi_1, \ldots, \pi_r$ (and $c_i \vee c_j$ has no more than k non zero terms). Then $c_i \vee c_j$ will contain $\pi_1, \ldots, \pi_{r+1}$. Repeat until $r = k$. Then repeat the process considering the probabilities $\pi_{k+1}, \ldots, \pi_n$

By reiterating these steps we will reach a matrix J' with columns $c'_1, \ldots, c'_n$ such that each element of column c'_i has higher probability than all elements of column c'_{i+1}. This is exactly the solution given by the greedy algorithm (modulo column permutations), i.e. $J' = J_k$.

Also it is immediate to see that no aligning and $\wedge, \vee$ operations will change the values in any column in J_k, i.e. we have reached a fixed point for the posterior entropy, i.e. the solution from the greedy algorithm is optimal. ■

C. An application to anonymity

Consider the following problem related to k-anonymity: design a deterministic query system where in order to conceal the real query from a (possibly malicious) server, the user embeds it in to a set of k queries: the server provides the response to all of the queries and the client retrieves the real one. In our setting, this corresponds to each observable having a pre-image of size exactly k. Let n be the number of possible queries. Considering the scenario that n is divisible by k, one of the solutions of this problem is the one suggested by Algorithm 1, relating to the k-hypergraph problem.

Given n secrets, there are $n!(\frac{n}{k}!)^{-1}(k!)^{-\frac{n}{k}}$ possible ways to satisfy these anonymity constraints. This means that there are $\approx 7 \times 10^{85}$ possible solutions when $n = 100$ and $k = 10$, and $\approx 4 \times 10^{19704}$ when $n = 10000$ and $k = 100$. How does the solution from the greedy algorithm in Section IV compare against other possible anonymity solutions? We consider Min and Shannon posterior entropies and the following three anonymity solutions: 1) the one from the greedy algorithm, 2) a random solution and 3) an un-optimal solution where the secrets with the highest probabilities, instead of being grouped in the first bin, are distributed in the other bins. For example, for 6 secrets and $k = 2$, the greedy solution would be $\{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}\}$ whereas the un-optimal solution 3) would be $\{\{x_1, x_4\}, \{x_2, x_5\}, \{x_3, x_6\}\}$.

The difference between these solutions can be very substantial. Figure 3 shows the values when the distribution over the input set is a binomial distribution, with parameter $p = 0.5$. In fact, it is possible to build a probability distribution for which the leakage gap between the optimal solution given by the greedy algorithm and the un-optimal can be made arbitrary large. Crucially Theorem 5 guarantees that the greedy solution is optimal for all leakage-supermodular entropies.

V. Probabilisitic Channels and Their Behaviour in Multiple Runs

The probabilistic channel design problem, as studied in [21], can be defined as follows

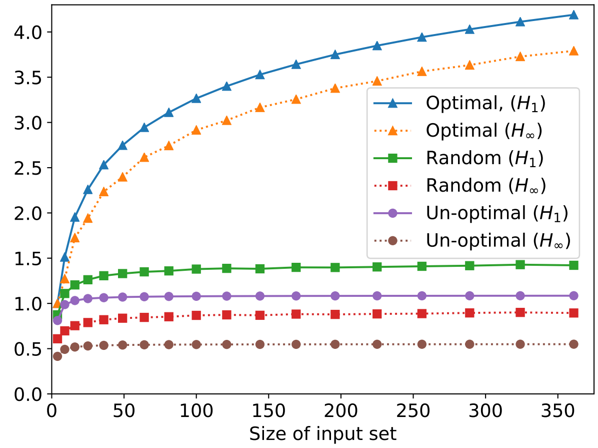

Fig. 3. Values of posterior Shannon and min-entropy for each of the anonymity solutions. The value of k is $\sqrt{n}$, where n is the size of the input set. For the random solution, the values plotted are the mean of 1000 samples.

Definition 9: The *probabilistic channel design problem* is to find $C \in \Gamma_\Omega$ that maximizes

$$\sum_{y \in \mathcal{Y}^+} p(y) F(\vec{p}_{\mathcal{X}|y})$$

constrained by $\mathbb{E}_{\pi,C}[u] \geq u_{min}$.

To illustrate the probabilistic channel design problem, consider the following example.

Example 2: in "timing attacks" [23] the adversary can obtain information about the secret by measuring a program execution times. It is possible to completely nullify this information leakage by forcing the system to always finish in a constant time (e.g. by adding dummy operations). Such strategy, however, might lead to an unacceptable decrease in overall performance. One might achieve a decrease in leakage while preserving better performance by employing "bucketing" [25]. Bucketing consists of selecting a small finite set of possible times for the system to finish its execution. Each secret value is then mapped to one of these times (buckets), decreasing leakage without excessively deteriorating the performance of the system.

As a toy-example, we consider the following probabilistic channel design problem for bucketing against a simplified timing attack introduced in [20]: Let $\mathcal{X} = \{x_1, x_2, x_3\}$ be a set of processes, the running time of x_i being i seconds. In each execution, the system runs only one process, and the goal of the adversary is to discover which one it is. Since the adversary can only observe the running time of the system, the output set can be represented as $\mathcal{Y} = \{1, 2, 3\}$. The hard constraints are $\Omega = \{(x_i, y) \in \mathcal{X} \times \mathcal{Y} \mid i \leq y\}$, as the time observed by the adversary can never be less than the execution time of the secret process. The utility function is given by $u(x_i, y) = i - y$, that is, the negative of the delay when process x_i is chosen and the system finishes the execution at time y; and take $u_{min} = -2/3$. A solution for $H = H_\infty$ and $\vec{\pi} = (1/2, 1/3, 1/6)$ is given by channel T in Fig. 4 (Left). Notice that $H_\infty(\pi, T) = H_\infty(\pi)$, and therefore the channel T does not leak any information for this choice of entropy measure and prior distribution.

T	2	3
x_1	$2/3$	$1/3$
x_2	1	0
x_3	0	1

T_1	2	3
x_1	$3/4$	$1/4$
x_2	$7/8$	$1/8$
x_3	0	1

T_2	2	3
x_1	$4/5$	$1/5$
x_2	$4/5$	$1/5$
x_3	0	1

Fig. 4. Left: Solution to the design problem of Example 2. Center and right: Channels T_1 and T_2 that respect the constraints of Example 2.

A. Multiple Runs of Probabilistic Channels

The solution to the problem in Definition 9 yields a channel that leaks the least information in a single execution, or when the secret is renewed for each execution in an i.i.d. manner. In many scenarios of interest, however, the adversary can observe multiple executions of a system while the value of the secret remains the same. Such is the case, for example, in timing attacks against cryptographic keys that remain unaltered for a large number of messages. The information leakage under these assumptions has been previously studied in QIF, specially regarding min-entropy [7], [12], [32].

After n executions of a system C, the adversary will have observed a sequence of outputs $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$. As the runs are independent and the secret value $x \in \mathcal{X}$ remains fixed, the probability of $\mathbf{y}$ being observed in n runs given x is exactly $C(x, y_1)C(x, y_2) \ldots C(x, y_n)$. We can thus model n executions of a system C by the channel C^n defined as follows:

Definition 10: Let $C : \mathcal{X} \to \mathcal{Y}$ and $n \in \mathbb{N}^*$. We define the channel $C^n : \mathcal{X} \to \mathcal{Y}^n$ as

$$C^n(x, (y_1, \ldots, y_n)) = \prod_{i=1}^{n} C(x, y_i).$$

Notice that executing a deterministic system with a fixed secret more than once does not leak any extra information, as the produced observables are all identical. Indeed, if D is a deterministic channel, $D^n =_\circ D$ for any $n > 0$.

A solution for a design problem as in Definition 9 will not, in general, be an optimal choice when repeated runs are taken into account, as it might be the case that $H(\pi, C_1) > H(\pi, C_2)$, but $H(\pi, C_1^n) < H(\pi, C_2^n)$ for some $n > 1$.

To illustrate this, we refer back to the design problem described in Example 2, whose optimal channel for single-run was T provided in Fig. 4 (Left). Consider the channels T_1 and T_2 in Figure 4, which respect the constraints of the problem but that are not solutions (i.e., do not maximize posterior min-entropy). Table I shows that, even though T_1 and T_2 are sub-optimal for a single execution, both are better choices than T whenever the system is run more than once. Specifically, T_1 is a better choice for two executions, and T_2 is more secure for three executions.

TABLE I
POSTERIOR MIN-ENTROPY OF T, T_1, T_2 WITH NUMBER OF EXECUTIONS

Number of runs	1	2	3
T	1	0.47	0.26
T_1	0.88	0.65	0.57
T_2	0.82	0.63	0.59

B. Deterministic Channels as a Solution for a Large Number of Runs

The discussion above reveals a shortcoming of the probabilistic channel design problem. If one is designing a system that can be run multiple number of times, using the solution of the problem as in Definition 9 could be suboptimal. In fact, the optimal solution of the probabilistic design problem could be much less secure than other feasible channels even for a small number of runs. For example, $H_\infty(\pi, T^{10}) \approx 0.01$, while $H_\infty(\pi, T^{10}) \approx 0.585$.

In this section, the problem considered is that of designing a channel when the number of executions is large. First, we consider the behaviour of a channel C^n when n tends to infinity. For each $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, let $N(y, \mathbf{y})$ denote the number of occurrences of $y \in \mathcal{Y}$ in $\mathbf{y}$, and let $t_{\mathbf{y}}(y) = {}^{N(y,\mathbf{y})}/_{n}$ be the *type* of $\mathbf{y}$, i.e., the relative frequency (or histogram) of the observable y in $\mathbf{y}$. Let $\sim_C$ to be the equivalence relation on $\mathcal{X}$ defined as follows:

$$x \sim_C x' \equiv \ \forall y,\ C(x, y) = C(x', y). \tag{5}$$

That is, these are all the x that have identical channel rows.

By the law of large numbers, as n grows, $t_{\mathbf{y}}$ approaches the probability $p_{\mathcal{Y}|x}(y) = C(x, y)$. The adversary can, therefore, deduce the probability distribution that is generating the string of observables. In other words, in limit, they are able to infer which equivalence class in $\mathcal{X}/{\sim_C}$ the secret belongs to. This suggests that for large enough n, the information revealed by C^n can be approximated by what is revealed by $C^\infty : \mathcal{X} \to \mathcal{X}/{\sim_C}$, defined as

$$C^\infty(x, S) = \begin{cases} 1, & \text{if } x \in S, \\ 0, & \text{otherwise.} \end{cases}$$

We call C^∞ the *limit channel* of C.

T^∞, T_1^∞	$[x_1]$	$[x_2]$	$[x_3]$
x_1	1	0	0
x_2	0	1	0
x_3	0	0	1

T_2^∞	$[x_1]$	$[x_3]$
x_1	1	0
x_2	1	0
x_3	0	1

Fig. 5. The limit channels for T, T_1 and T_2.

The above discussion is formalized in the following proposition. Its proof is similar to that of Theorem 1 in [6], which studied asymptotic leakage properties for min-entropy. We modify it to obtain a result applicable for all entropy measures.

Proposition 4: Let $C : \mathcal{X} \to \mathcal{Y}$ be a channel and let $C^\infty : \mathcal{X} \to \mathcal{X}/{\sim_C}$ be its limit channel. For all $n > 0$, $\pi \in \mathbb{D}\mathcal{X}$ and all entropy measures H, we have

$$H(\pi, C^n) \geq H(\pi, C^\infty), \tag{6}$$

$$\lim_{n\to\infty} H(\pi, C^n) = H(\pi, C^\infty). \tag{7}$$

The proof of Proposition 4 is provided in Appendix-B.

Proposition 4 shows that all channels behave like a deterministic channel after a large enough number of repeated executions with a fixed input, but the rate of convergence depends on the choice of entropy measure. To illustrate this, consider again channel T from Example 2. Because all its rows are different, T^∞ completely reveals the secret, and therefore the posterior Rényi entropy $H_\alpha(\pi, T^\infty)$ will be 0 for any α.

Figure 6 shows the values of $H_\alpha(\pi, T^i)$ for some choices of α, including the limit cases of min-entropy and Shannon entropy. Notice that, while all values tend to $H_\alpha(\pi, T^\infty) = 0$, they do so at different rates. The graph in Figure 7 further illustrates this difference, presenting the smallest value of i for which $H_\alpha(\pi, T^i) < 0.05$, for varying values of α.

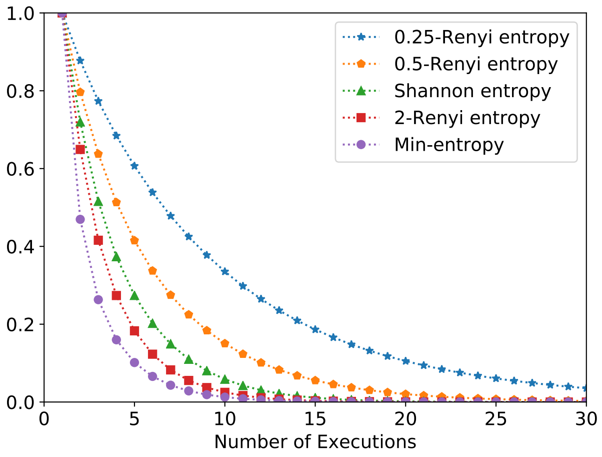

Fig. 6. Values of posterior Rényi entropies of different orders, for increasing number of executions of channel T.

Being a channel with output set $\mathcal{X}/{\sim_C}$, a limit channel C^∞ is *not* in the solution space of the channel design problem. In our next result, we establish that there is always a deterministic channel $D \in \Gamma_\Omega$ that satisfies $\mathbb{E}_{\pi,D}[u] \geq u_{min}$ such that $C^\infty \sqsupseteq_\circ D$. As D is deterministic, $D^n =_\circ D$ for all n. Therefore, given any channel C that satisfies the constraints of the problem, it is possible to find a deterministic channel D that also satisfies those constraints and asymptotically leaks as little information as C.

Proposition 5: Given a probabilistic channel design problem as in Definition 9, let C be a a channel such that $C \in \Gamma_\Omega$ and $\mathbb{E}_{\pi,C}[u] > u_{min}$. Then, there exists a deterministic channel $D \in \Gamma_\Omega$ such that $\mathbb{E}_{\pi,D}[u] > u_{min}$ and $C^\infty \sqsupseteq_\circ D$.

Proof: Given a channel C, let the deterministic channel D be obtained by mapping all the secrets of each equivalence class $S \in \mathcal{X}/{\sim_C}$ to one output $y_S \in \mathcal{Y}$ allowed by the hard constraints – i.e., y_S is an element of the set $\mathcal{Y}_S = \{y \in$

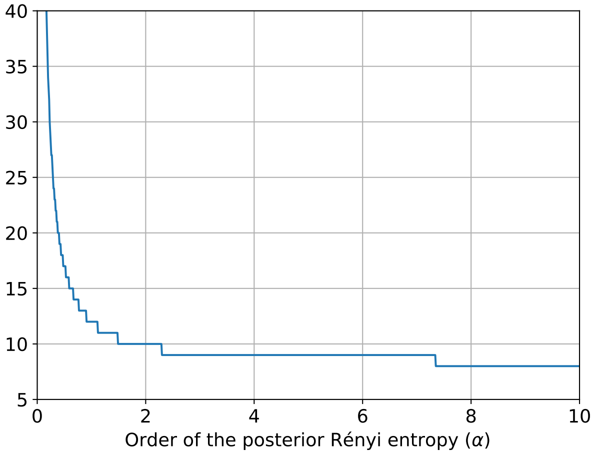

Fig. 7. The smallest value of i for which $H_\alpha(\pi, T^i) \leq 0.05$, for varying values of α.

$\mathcal{Y} \mid \forall x \in \mathcal{S}, (x, y) \in \Omega\}$. We can guarantee that $E_{\pi,D}[u] \geq u_{min}$ by choosing y_S that maximizes the value of the utility function on that equivalence class, namely

$$y_S \in \arg\max_{y \in \mathcal{Y}_S} \sum_{x \in \mathcal{S}} \pi(x) u(x, y).$$

Notice that, as all the secrets on each equivalence class get mapped to the same output, the procedure above does not divide equivalence classes, although it might join them together – that is, $\sim_C$ is a finer relation than $\sim_D$. As D is deterministic, this implies that $C^\infty \sqsupseteq_\circ D$. Explicitly, $D = C^\infty R$, where $R : \mathcal{X}/{\sim_C} \to \mathcal{Y}$ is the channel given by $R(S, y) = 1$ if $y = y_S$ and 0 otherwise. ■

Given any channel C that respects the constraints of a channel design problem, the proof of Proposition 5 provides an algorithm to find a deterministic channel D that leaks no more than C in the limit case. Consider, for example, channels T, T_1 and T_2 in Figure 4. This algorithm yields the following channels D_1 (for T and T_1) and D_2 (for T_2).

D_1	1	2	3
x_1	1	0	0
x_2	0	1	0
x_3	0	0	1

D_2	2	3
x_1	1	0
x_2	1	0
x_3	0	1

Comparing with the limit channels in , Figure 5, we can see that $D_1 =_\circ T^\infty$, T_1^∞ and $D_2 =_\circ T_2^\infty$.

A immediate consequence of Proposition 5 is that the solution of the deterministic channel design problem is also an asymptotically optimal solution for the probabilistic channel design problem under a large number of executions. Take D_2 for example, which is the optimal deterministic channel for the constraints of Example 2. We have $H_\infty(\pi, D_2) \approx 0.585$, and from Table I we can infer that for min-entropy, D_2 leaks less information than T^i for any $i \geq 2$, and than T_1^i for any $i \geq 3$.

VI. Conclusion and Future Work

The channel design problem for general constraints has only recently started to be explored in the literature [20]–[22]. This work investigates the problem of designing a deterministic system whose information leakage is minimal.

The results proven apply to the generalized family of entropies defined in Section II-B, which subsumes most entropy measures in the literature. A first result is the NP-completeness of the deterministic channel design problem, and that, in general, there is no solution that is optimal for all choices of entropies.

These negative results prompted the exploration of the k-complete hypergraph problem, for which a computationally efficient solution is achievable that is optimal for most of the commonly used entropy measures in the literature. In proving the optimality of that solution, leakage-supermodularity was introduced. This concept may prove beneficial for future research in QIF, exploring the connection between common entropy measures and supermodular functions established by Theorem 2.

Other possible future directions of research include exploring different classes of channel design problems, for which one can find efficient or universal solutions, and applications of this framework to real-life systems.

References

[1] P. Ah-Fat and M. Huth. Optimal accuracy-privacy trade-off for secure computations. *IEEE Transactions on Information Theory*, pages 1–1, 2018.

[2] Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, and Catuscia Palamidessi. Differential Privacy: on the trade-off between Utility and Information Leakage. In *Postproceedings of the 8th Int. Worshop on Formal Aspects in Security and Trust (FAST)*, volume 7140 of *LNCS*, pages 39–54. Springer, 2011.

[3] Mário S. Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. Axioms for information leakage. In *Proc. of CSF*, pages 77–92, 2016.

[4] Mário S. Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pages 265–279, 2012.

[5] S Arimoto. Information measures and capacity of order α for discrete memoryless channels. *Topics in information theory*, 1977.

[6] Michele Boreale, Francesca Pampaloni, and Michela Paolini. Asymptotic information leakage under one-try attacks. In *Proc. of FOSSACS*, volume 6604 of *LNCS*, pages 396–410. Springer, 2011.

[7] Michele Boreale, Francesca Pampaloni, and Michela Paolini. Asymptotic information leakage under one-try attacks. *Mathematical Structures in Computer Science*, 25(2):292–319, 2015.

[8] David Clark, Sebastian Hunt, and Pasquale Malacaria. Quantitative information flow, relations and polymorphic types. *J. of Logic and Computation*, 18(2):181–199, 2005.

[9] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. J. Wiley & Sons, Inc., second edition, 2006.

[10] Cynthia Dwork. Differential privacy. In *Proc. of ICALP*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.

[11] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theor. Comp. Sci.*, 9(3–4):211–407, 2014.

[12] Barbara Espinoza and Geoffrey Smith. Min-entropy as a resource. *Inf. and Comp.*, 226:57–75, 2013.

[13] Xun Gong and Negar Kiyavash. Quantifying the information leakage in timing side channels in deterministic work-conserving schedulers. *IEEE/ACM Transactions on Networking*, 24(3):1841–1852, 2016.

[14] Masahito Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Transactions on Information Theory*, 57(6):3989–4001, 2011.

[15] Jonathan Heusser and Pasquale Malacaria. Quantifying information leaks in software. In *Twenty-Sixth Annual Computer Security Applications Conference, ACSAC 2010, Austin, Texas, USA, 6-10 December 2010*, pages 261–269, 2010.

[16] Mitsugu Iwamoto and Junji Shikata. Revisiting conditional rényi entropies and generalizing shannons bounds in information theoretically secure encryption. Technical report, Cryptology ePrint Archive 440/2013, 2013.

[17] Mitsugu Iwamoto and Junji Shikata. Information theoretic security for encryption based on conditional rényi entropies. In Carles Padró, editor, *Information Theoretic Security*, pages 103–121, Cham, 2014. Springer International Publishing.

[18] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer US, Boston, MA, 1972.

[19] Ali Khoshgozaran and Cyrus Shahabi. Private information retrieval techniques for enabling location privacy in location-based services. In *Privacy in Location-Based Applications*, volume 5599, pages 59–83. Springer, 2009.

[20] MHR Khouzani and Pasquale Malacaria. Leakage-Minimal Design: Universality, Limitations, and Applications. In *30th IEEE Computer Security Foundations Symposium (CSF)*, pages 305–317. IEEE, 2017.

[21] MHR Khouzani and Pasquale Malacaria. Generalised Entropies and Metric-Invariant Optimal Countermeasures for Information Leakage under Symmetric Constraints. *IEEE Transactions on Information Theory*, 2018.

[22] MHR. Khouzani and Pasquale Malacaria. Optimal channel design: A game theoretical analysis. *Entropy*, 20(9), 2018.

[23] Paul C. Kocher. Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems. In Neal Koblitz, editor, *Advances in Cryptology — CRYPTO '96*, pages 104–113, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

[24] Boris Köpf and David A. Basin. An information-theoretic model for adaptive side-channel attacks. In *Proc. of CCS*, pages 286–296. ACM, 2007.

[25] Boris Köpf and Geoffrey Smith. Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks. In *Proc. of CSF*, pages 44–56. IEEE, 2010.

[26] Albert W Marshall, Ingram Olkin, and Barry C Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Mathematics In Science And Engineering, Academic Press, 1979.

[27] Massey. Guessing and entropy. In *Proceedings of the IEEE Int. Symposium on Information Theory*, page 204. IEEE, 1994.

[28] Corina S. Pasareanu, Quoc-Sang Phan, and Pasquale Malacaria. Multi-run side-channel analysis using symbolic execution and max-smt. In *IEEE 29th Computer Security Foundations Symposium, CSF 2016, Lisbon, Portugal, June 27 - July 1, 2016*, pages 387–400, 2016.

[29] Alfréd Rényi. On Measures of Entropy and Information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, pages 547–561, 1961.

[30] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 625–56, 1948.

[31] B.D. Sharma and D.P. Mittal. New non-additive measures of entropy for discrete probability distributions. *Journal of Mathematical Science (Soc. Math. Sci., Calcutta, India*, 10:28–40, 1975.

[32] D. M. Smith and G. Smith. Tight bounds on information leakage from repeated independent runs. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 318–327, Aug 2017.

[33] Geoffrey Smith. On the foundations of quantitative information flow. In *Proc. of FOSSACS*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.

[34] George Theodorakopoulos, Reza Shokri, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services. In *Proc. of WPES*. ACM, 2014.

[35] Constantino Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1-2):479–487, 1988.

Appendix A
Alternative proof for Theorem 5

Let H be a leakage-supermodular entropy and π be a prior distribution given by the probability vector $\vec{\pi} = (\pi_1, \dots, \pi_n)$. Without loss of generality, we assume $\pi_i \geq \pi_j$ whenever $i < j$. Let C_k be the solution yielded by Algorithm 1.

Let $m = \lceil n/k \rceil - 1$. From Proposition 2, and from symmetry and expansibility of G_F, we obtain

$$H(\pi, C_k) = \eta \left(\sum_{i=0}^{m} G_F(\pi_{ik}, \pi_{ik+1}, \dots \pi_{(i+1)k}) \right), \tag{8}$$

in which we define $\pi_j = 0$ if $j > n$.

Let C be any channel that respects the deterministic channel design constraints, and hence the preimage of each output consists of at most k inputs. For each $y \in f_C(\mathcal{X})$, let $\mathbf{c}_y$ be the vector formed by the elements of the set $\{\pi_i \mid x_i \in f_C^{-1}(y)\}$, followed by $k - |f_C^{-1}(y)|$ zeros. From proposition 2 and expansibility of G_F, we obtain

$$H(\pi, C) = \eta \left(\sum_{y \in f(\mathcal{X})} G_F(\mathbf{c}_y) \right). \tag{9}$$

For each $i \leq k$, let y_i be such that π_i is an entry of $\mathbf{c}_{y_i}$. We permutate the elements of each $\mathbf{c}_{y_i}$ so that π_i is in its ith entry. As G_F is symmetric, this permutation does not change the value of (9).

Notice that, if we substituted $\mathbf{c}_{y_1}$ by $\mathbf{c}_{y_1} \vee \mathbf{c}_{y_2}$ and $\mathbf{c}_{y_2}$ by $\mathbf{c}_{y_1} \wedge \mathbf{c}_{y_2}$, both π_1 and π_2 would be entries of $\mathbf{c}_{y_1}$. Moreover, supermodularity of G_F implies that the value of (9) would not decrease. By repeating the process with $\mathbf{c}_{y_1}$ and $\mathbf{c}_{y_3}, \dots, \mathbf{c}_n$, $\mathbf{c}_{y_1}$ would contain $\pi_1, \dots, \pi_k$ and the value of (9) would not decrease.

More formally, we define

$$\mathbf{c}_y^{(1)} = \begin{cases} \mathbf{c}_{y_1} \vee \mathbf{c}_{y_2} \vee \cdots \vee \mathbf{c}_{y_k} & \text{if } y = y_1, \\ (\mathbf{c}_{y_1} \vee \cdots \vee \mathbf{c}_{y_{i-1}}) \wedge \mathbf{c}_{y_i} & \text{if } y = y_i, i > 1, \\ & \text{and } \forall j < i, y_i \neq y_j \\ \mathbf{c}_y & \text{otherwise.} \end{cases}$$

By the previous discussion, $\mathbf{c}_{y_1}^{(1)} = (\pi_1, \dots, \pi_k)$. As G_F is supermodular and η is increasing, we obtain

$$\eta \left(\sum_{y \in f_C(\mathcal{X})} \mathbf{c}_y^{(1)} \right) \geq H(\pi, C).$$

We repeat the procedure, now considering $\pi_{k+1}, \dots, \pi_{2k}$. For $k < i \leq 2k$, let y_i be the output such that π_i is an entry $\mathbf{c}_{y_i}^{(1)}$ (notice that $y_1 \neq y_i$ for all $k < i \leq 2k$). Finally, we permutate the vectors such that π sits at the $(i-k)$th entry, and define

$$\mathbf{c}_y^{(2)} = \begin{cases} \mathbf{c}_{y_{k+1}}^{(1)} \vee \mathbf{c}_{y_{k+2}}^{(1)} \vee \cdots \vee \mathbf{c}_{y_{2k}}^{(1)} & \text{if } y = y_{k+1}, \\ (\mathbf{c}_{y_{k+1}}^{(1)} \vee \cdots \vee \mathbf{c}_{y_{i-1}}^{(1)}) \wedge \mathbf{c}_{y_i}^{(1)} & \text{if } y = y_i, i > k+1, \\ & \text{and } \forall j < i, y_i \neq y_j, \\ \mathbf{c}_y^{(1)} & \text{otherwise.} \end{cases}$$

Hence, $\mathbf{c}_{y_1}^{(2)} = (\pi_1, \dots, \pi_k)$ and $\mathbf{c}_{y_{k+1}}^{(2)} = (\pi_{k+1}, \dots, \pi_{2k})$. Moreover, supermodularity of G_F $\eta \left(\sum_y \mathbf{c}_y^{(2)} \right) \geq H(\pi, C)$.

Recall that $m = \lceil n/k \rceil - 1$. Proceeding in this manner, we obtain vectors $\mathbf{c}_y^m$ that are either of the form $(\pi_{ik+1}, \ldots, \pi_{(i+1)k})$ for some $i \leq m$, or completely formed by zeros. As $G_F(0, \ldots, 0) = 0$, Equation (8) yields

$$H(\pi, C_k) = \eta \left(\sum_{y \in f_C(\mathcal{X})} \mathbf{c}_y^{(m)} \right) \geq H(\pi, C).$$

Therefore, the posterior entropy induced by C is never greater than the one induced by C_k.

Appendix B
Proof of Proposition 4

Proof: (6): Consider the channel $E : \mathcal{X}/{\sim_C} \to \mathcal{Y}$ defined as $E([x], y) = C(x, y)$. Then, $C^n = C^\infty E^n$, which implies $C^\infty \sqsupseteq_\circ C^n$, and the result follows from Theorem 1.

(7): For each $S \in \mathcal{X}/{\sim_C}$, let p_S be the channel row of that class. That is, let $p_S \in \mathbb{D}\mathcal{Y}$ be defined as $p_S(y) = C(x, y)$, for some $x \in S$. Moreover, given an $\epsilon > 0$, let:

$$U_S^n(\epsilon) = \{\mathbf{y} \in \mathcal{Y}^n \,|\, D_{KL}(t_\mathbf{y} \,||\, p_S) \leq \epsilon\}$$

where D_{KL} is the Kullback-Leibler divergence. Intuitively, $U_S^n(\epsilon)$ is a set of sequences of $\mathcal{Y}^n$ that are "close" to p_S. Then, Theorem 11.2.1 of [9] yields that for all $S \in \mathcal{X}/{\sim_C}$,

$$p_S^n\left(U_S^n(\epsilon)\right) \geq r(n) \tag{10}$$

where $r(n) = 1 - (n+1)^{|\mathcal{Y}|} 2^{-n\epsilon}$, and $p_S^n \in \mathbb{D}(\mathcal{Y}^n)$ is given by $p_S^n(y_1, \ldots, y_n) = p_S(y_1) p_S(y_2) \ldots p_S(y_n)$. As n gets larger, $r(n)$ tends to 1, and therefore so does $p_S^n\left(U_S^n(\epsilon)\right)$.

For each $\mathbf{y} \in \mathcal{Y}^n$, we pick $S_\mathbf{y} \in \mathcal{X}/{\sim_C}$ such that $\exists x \in S_\mathbf{y}, \forall x' \in \mathcal{X},\ \pi(x) C^n(x, \mathbf{y}) \geq \pi(x') C^n(x', \mathbf{y})$. We define $G_n : \mathcal{Y}^n \to \mathcal{X}/{\sim_C}$ as

$$G_n(\mathbf{y}, S) = \begin{cases} 1, & \text{if } S = S_\mathbf{y}, \\ 0, & \text{otherwise.} \end{cases}$$

That is, G_n is a channel that maps each sequence of outputs to the equivalence class of the secret value that most probably produces it. The idea of the proof is that, by considering the post-processing of C^n by G_n, we can use Theorem 1 to bound $H(\pi, C^n)$ from above by $H(\pi, C^n G_n)$.

Define $A_S^n = \{\mathbf{y} \in \mathcal{Y}^n \,|\, S = S_\mathbf{y}\}$. By definition of G_n, we have

$$(C^n G_n)(x, [x]) = \sum_{\mathbf{y} \in \mathcal{Y}^n} C^n(x, \mathbf{y}) G_n(\mathbf{y}, [x]) = \sum_{\mathbf{y} \in A_{[x]}^n} C^n(x, \mathbf{y})$$

By Lemma 3 in [7], there exists $\epsilon_1 > 0$ such that $U_S^n(\epsilon_1) \subset A_S^n$ for all large enough n. Thus, for such n,

$$(C^n G_n)(x, [x]) = \sum_{\mathbf{y} \in A_{[x]}^n} C^n(x, \mathbf{y}) \geq \sum_{\mathbf{y} \in U_{[x]}^n(\epsilon_1)} C^n(x, \mathbf{y}).$$

By definition, the rightmost summation is equal to $p_S^n\left(U_{[x]}^n(\epsilon_1)\right)$. Hence, from (10) we obtain, for all $x \in \mathcal{X}$,

$$(C^n G_n)(x, [x]) \geq r(n).$$

That is, $(C^n G_n)(x, S) \geq r(n)$ whenever $C^\infty(x, S) = 1$. As all rows of $C^n G_n$ sum to 1, $(C^n G_n)(x, S) \leq 1 - r(n)$ whenever $C^\infty(x, S) = 0$. Hence, reasoning about $C^n G_n$ and C^∞ as vectors of $\mathbb{R}^{|\mathcal{X}||\mathcal{X}/\sim_C|}$, the norm of their difference can be bounded as

$$\|C^n G_n - C^\infty\|_\infty \leq 1 - r(n).$$

Therefore, $\lim_{n\to\infty} C^n G_n = C^\infty$ and, by continuity, $\lim_{n\to\infty} H(\pi, C^n G_n) = H(\pi, C^\infty)$. Now, Theorem 1 and Equation (6) imply that, for all $n \in \mathbb{N}$

$$H(\pi, C^n G_n) \geq H(\pi, C^n) \geq H(\pi, C^\infty),$$

and hence (7) follows from the sandwich theorem. ■

As a visualization of the technique used in the proof of Proposition 4, some channels of the sequence $T_1^n G_n$ are presented below, where T_1 is as in Figure 4 and $\vec{\pi} = (1/2, 1/3, 1/6)$. The channel entries are rounded to two decimal places. Notice that the sequence $T_1^n G_n$ converges to T_1^∞ in Figure 5.

$T_1^2 G_2$	$[x_1]$	$[x_3]$
x_1	0.94	0.06
x_2	0.98	0.02
x_3	0	1

$T_1^4 G_4$	$[x_1]$	$[x_2]$	$[x_3]$
x_1	0.68	0.32	0
x_2	0.41	0.59	0
x_3	0	0	1

$T_1^{20} G_{20}$	$[x_1]$	$[x_2]$	$[x_3]$
x_1	0.77	0.23	0
x_2	0.23	0.77	0
x_3	0	0	1

$T_1^{100} G_{100}$	$[x_1]$	$[x_2]$	$[x_3]$
x_1	0.96	0.04	0
x_2	0.07	0.93	0
x_3	0	0	1