Relative Perfect Secrecy: Universally Optimal Strategies and Channel Design

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

***Abstract*—Perfect secrecy describes cases where an adversary cannot learn anything about the secret beyond its prior distribution. A classical result by Shannon shows that a necessary condition for perfect secrecy is that the adversary should not be able to eliminate any of the possible secrets. In this paper we answer the following fundamental question: What is the lowest leakage of information that can be achieved when some of the secrets have to be eliminated? We address this question by deriving the minimum leakage in closed-form, and explicitly providing "universally optimal" randomized strategies, in the sense that they guarantee the minimum leakage irrespective of the measure of entropy used to quantify the leakage. We then introduce a generalization of Rényi family of asymmetric measures of leakage which generalizes the g- leakage and show that a slight modification of our strategies are optimal with respect to an important class of such measures. Subsequently, we show that our schemes constitute the Nash Equilibria of closely related two-person zero sum games. This game perspective provides implicit solutions for a wider set of structural constraints and asymmetric entropies. Finally we demonstrate how this work can also be seen as designing a universally optimal channel given a specified prior.**

I. Introduction

It is increasingly accepted that in many setups, like side channels and database queries, the quantification of leakage of confidential information is essential. In recent years considerable progress has been made in the field of quantitative information flow both on the theory and applications, e.g. [1]–[4]. Particularly important for the field are advances on fundamental security guarantees of leakage measures (what security can be achieved with that measure) and robust techniques and results (how much a technique or result is valid across different notions of leakage).

This work contributes to both leakage guarantees and robustness in that it investigates channels which are optimal, in the sense of guaranteeing minimum leakage, for all possible notions of leakage. By channel here we mean a probabilistic system where for each secret, there is a probability distribution over the behaviors or states that can be observed by an attacker.

In its essence the aim of this paper is to investigate the concept of *relative perfect secrecy*, a concept inspired by Shannon perfect secrecy theorem [5]. Following Shannon's theorem a "perfect" encryption scheme must in part be such that an adversary cannot eliminate any secret key by observing the system. We study here the case where perfect secrecy is not possible, in particular, when the operational constraints on the system are such that each observation allows the attacker to eliminate a certain number of possible secrets. As real world examples of such constraints we can think of timing observations which allow the attacker to eliminate certain keys, or Geo-location privacy where some geographic locations are impossible given the observations, or in private querying when not the entire database but a portion of it is downloaded, or when some but not all features of a device/browser can be suppressed against fingerprinting. To formalize the problem we think of N possible secrets with a given distribution, and that an observation on the system allows elimination of $N - k$ possible secrets. We then ask the question: Is there an "optimal" channel given these constraints, i.e., a channel with minimum leakage?

A challenging problem in investigating optimal channels is that there are several notions of information leakage, e.g., Shannon based [6], Min Entropy based [7], Bayesian [8] and g-leakage [9] and they have largely incomparable behavior. So there is no a priori reason why the notion of optimality should be robust. Moreover the few robustness results in the field have been hard to prove (e.g. the proof of the Coriaceous Conjecture [10]). The universal optimality results in this paper are hence non-trivial both in their meaning and in their proof, and they extend the toolkit of robust methods and results in the field.

Our presentation uses the notion of *cloaks*. A cloaking scheme is a non-cryptographic method of hiding secrets by conflating them with larger sets of possibilities, potentially employing randomization. In this context perfect secrecy can only be achieved when the cloaks are the entire secret space.

A simple example of cloaking comes from the famous Monty Hall problem, where there are multiple closed doors behind one of which is the prize, i.e., the secret. The player makes an initial guess, after which the game host has to open one door, but it should not be the one that the player has chosen and also not the prize-holding door, thus eliminating one possible secret. The cloak is then the set of the remaining closed doors, since each can harbor the secret. The game host acts as a channel. The famously tricky

question[1] of "should the player revise her initial guess?" can be equivalently asked as: "does this channel leak?". The fact that the player should change their initial guess is indeed because the channel leaks information.

The focus of this work is foundational. In particular, the list of our contributions is as follows:

***Road-map and Contributions**:* We formalize the problem of minimizing the information leakage given a prior distribution of the secret and a cap on the size of the cloaks, where the information leakage is quantified as a difference between the prior and posterior uncertainties of an adversary for a generic entropy measure (Section II). In Section III, we express and prove our main result (Theorem 1), that is, we provide the lowest achievable leakage across all (potentially probabilistic) cloaking strategies in closed form. We explicitly construct randomized strategies that achieve this information theoretical bound, and establish that they are universally optimal, in that they achieve minimum leakage with respect to *any* choice of entropy measure that satisfies three mild conditions: *core-concavity* (which we define in the text), symmetry and expansiblity. Next, in Section IV, we consider non-symmetric (gain-based) entropies, introduce a generalization of g-entropy and g-leakage, and establish a natural extension of our main result to this class of entropies (Theorem 2). In Section V, we make a connection between designing minimum leakage channels and 2-player-zero-sum games with respect to g-leakage, which enables us to give a Linear Program that produces the optimal strategies for any secret-dependent cloaking constraints and any gain function g. Finally, in Section VI, we numerically investigate the effect of the maximum allowable size of the cloaks, the choice of the entropy, comparing with baseline of uniform randomization, and checking the effect of the knowledge of the prior distribution.

Our proofs follow non-trivial techniques that we believe will add to the theoretical toolbox of the research community. Despite the theoretical nature of this work, we envisage possible applications of our results in fields such as side channels countermeasures in the style of bucketing [11], [12], in privacy preserving mechanisms like crowd-based anonymity protocols [13], (Geo)-location privacy [14], [15], or obfuscation-based web searching [16], etc. Detailed investigation of these connections and potential practical implementations will be part of our future work.

***Related Literature**:* The measures of leakage this work refers to have been the object of many works in the past decade. These works mostly use Shannon (e.g. [6]), Min Entropy (e.g. [7]) and Bayes risk [8]. More recent works [4], [9] introduce the notion of g-leakage in order to model leakage scenarios that are not satisfactorily modeled using standard entropic measures of leakage. Our results apply to all these approaches.

[1]It is said that Paul Erdős at first thought the player should not switch.

The leakage ordering has been the object of robustness analysis in several works e.g. [10], [17]. These works are quite different in nature from our work: here we design a channel given a prior on the secret, there the channel is given and the ordering is over all possible priors.

Information-theoretic relaxation of the absolute-privacy through cloaking is also investigated in [18] in the context of Internet search engines. The paper however does not quantify the leakage and provides only a lower-bound on the posterior entropy. Also, the proposed policies are heuristic and no claim or quantification of optimality is made.

There has been little work relating quantitative information flow and game theory. Few exceptions are [19], [20]. In [19] optimal protection against timing attacks is proven as an equilibrium in a game between the defender and the adversary. There optimal refers to crypto optimal not information theoretical optimal. [20] formalizes the privacy-vs-utility trade-offs in the context of Geo-location privacy as a game of inference against an attacker, and presents an implicit framework based on LP for derivation of the Stackelberg strategies. It does not however provide the strategies in closed-form nor does it consider robustness. In general there has been little work on quantifying information flow in interactive systems and strategic behavior. Some notable exceptions are the works of [21]–[23]. None of these works however is game theoretical in nature nor seems to fit the context of channel design as the channel is typically a given.

Non-cryptographic schemes are also investigated in the context of secret sharing and key-distribution [24], but in such scenarios, the goal is to hide a secret from a third party (e.g. eavesdropper), while in our setting, there is only one system vs. an adversary, and no "communication" of information is intended.

II. Model

Let the random variable $\boldsymbol{\theta}$ represent the secret. It can take one of the n possibilities from the (discrete finite) set $\Theta := \{\theta_1, \ldots, \theta_n\}$ that are generated independently and identically distributed (i.i.d.) over time, according to the (categorical) distribution $\boldsymbol{P}$, that is, at each instance and irrespective of the past history, $\Pr(\boldsymbol{\theta} = \theta_i) = \boldsymbol{P}(\theta_i) := p_i$. Whenever not ambiguous, we only refer to a secret by its index, e.g., we may use i to designate θ_i. Without loss of generality, assume $\mathrm{supp}(\boldsymbol{P}) = \Theta$, i.e., $p_i > 0$, $\forall i$. Also, without loss of generality, assume p_is are in descending order, i.e., $p_1 \geq \ldots \geq p_n$.[2]

We make a worst-case (for the defender) assumption about the *adversaries*: that they know the true distribution

[2]Note that throughout the paper by terms such as *descending/decreasing*, we mean *non-ascending/non-increasing*, unless accompanied by the qualifier: *strictly*. The same applies to terms like *positive*, *concave*, distinguished from *strictly positive*, *strictly concave*, and so on.

according to which the secrets are generated.[3] That is, we take $\boldsymbol{P}$ to be the prior belief of the *adversary* about the secret, hence we will simply refer to $\boldsymbol{P}$ as *the prior*.[4]

The defender, observing the (realization of the) secret, chooses a *cloak* M to submit. A cloak is a subset of the secrets including the actual one. As we noted before, in the absence of any restriction on the choice of the cloak, the best trivial cloak is the entire secret space, but in practice, the choice of the cloak is restricted. Let the set of *permissible* cloaks given secret realization θ be denoted by $\mathcal{M}(\theta)$. The minimal requirement on any $M \in \mathcal{M}(\theta)$ is that $\theta \in M$: the cloak must include the secret itself. The action space of the defender is hence $\mathcal{M} := \cup_{\theta\in\Theta}\mathcal{M}(\theta)$. Up to Section V, we consider a *size-capped* cloaking constraint, that is, each cloak is restricted to have at most k elements, where k, $1 \leq k \leq n$, is a given parameter of the problem. Specifically, $\mathcal{M}(\theta) = \{M \subset \Theta : \theta \in M, |M| \leq k\}$, and $\mathcal{M} = \{M \subset \Theta : |M| \leq k\}$.

We will use the following toy example to clarify the concepts: Suppose the secret space is $\Theta = \{1, 2, 3\}$ with the prior distribution of $\boldsymbol{P} = (5/9, 3/9, 1/9)$, and the cloaks are limited to at most $k = 2$ elements. Then the set of feasible cloaks for secret 1 is $\mathcal{M}(1) = \{\{1\}, \{1, 2\}, \{1, 3\}\}$, and the set of all feasible cloaks is $\mathcal{M} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$.

A deterministic cloaking plan of the defender, or a *deterministic cloaking strategy*, denoted by $\boldsymbol{d}$, is a function from the set of secrets to set of permissible cloaks for that secret.[5] Specifically, the space of the deterministic cloaking strategies of the defender is $\mathcal{D} := \{\boldsymbol{d} : \Theta \to \mathcal{M} \text{ s.t. } \boldsymbol{d}(\theta) \in \mathcal{M}(\theta), \forall\theta \in \Theta\}$. An adversary (which could be the service provider itself) observes the cloak M and updates his belief about the distribution of the secret.

The defender may incorporate randomness in her cloaking strategy, as intuitively, this would increase the uncertainty of the adversary. A *randomized* (cloaking) strategy, which we designate by $\boldsymbol{\delta}$, assigns each secret a probability distribution over the set of permissible cloaks for that secret. Specifically, the space of randomized cloaking strategies is $\mathscr{D} := \{\boldsymbol{\delta} : \Theta \to \Delta\mathcal{M} \text{ s.t. } \forall\theta \in \Theta, \operatorname{supp}(\boldsymbol{\delta}(\theta)) \subseteq \mathcal{M}(\theta)\}$, where $\Delta\mathcal{M}$ represents the set of all probability distributions over $\mathcal{M}$. We use the notation $\boldsymbol{\delta}(M; \theta)$ to designate the probability at which, under the randomized strategy of $\boldsymbol{\delta}$, the defender chooses cloak M when her secret is θ. Using this notation, the space of randomized cloaking strategies can be specified by the following two conditions:

$$\boldsymbol{\delta}(M; \theta) \geq 0 \quad \text{for all } \theta \in \Theta,\ M \in \mathcal{M}; \tag{1a}$$

$$\sum_{M\in\mathcal{M}(\theta)} \boldsymbol{\delta}(M; \theta) = 1 \quad \text{for all } \theta \in \Theta. \tag{1b}$$

Referring to our toy example, an instance of a deterministic cloaking strategy is: $\{1 \to \{1, 2\}, 2 \to \{1, 2\}, 3 \to \{1, 3\}\}$, and an example of a randomized strategy can be: $\{1 \to (\{1, 2\} +_{1/2} \{1, 3\}), 2 \to \{1, 2\}, 3 \to \{1, 3\}\}$.[6] Clearly, any deterministic strategy can also be represented as a randomized strategy with degenerate distributions.

The objective of the defender is to receive the sensitively secret-dependent service while leaking the least information about the secret to the adversary. To quantify the expected leakage of information, we need to consider a measure of information content or *entropy* of the secret in the eye of the adversary, to be compared before and after the interaction. n in the entropy quantifies the leakage of information to the adversary.

Interpretation of Cloaking Strategies as Channels: A channel is defined as a triple (X, Y, C) where X and Y are finite sets. X is the set of secret input values and Y the set of observables. C is an $|X| \times |Y|$ matrix (the channel matrix) whose entries are between 0 and 1 and whose rows each sum to 1. The channel matrix represents the conditional probability of an observable given a secret: $C[x, y]$ is the probability of observing y given x.

In the context of this work we can associate to a cloaking strategy the channel defined as follows: X is the set of secrets Θ, Y is the set of cloaks $\mathcal{M}(\theta)$. The coefficients for the channel matrix C are

$$C[\theta, M] = \boldsymbol{\delta}(M; \theta)$$

For the inverse direction a channel can be interpreted as a cloaking strategy of size k by considering the observables' pre-images. In terms of the channel matrix the pre-image of an observable is the set of secrets θ s.t. $C[\theta, M] > 0$, i.e. the non-zero elements in the column corresponding to the observable. Given a channel we can associate a cloaking strategy of size k if:

1) the pre-image of any observable has no more than k elements;
2) no two observables have the same pre-image.

For a channel satisfying the above properties we can identify each observable with the cloak that is its pre-image.

Notice that in most works in Quantitative Information Flow the channel is given. The problem tackled in this paper is different: we are given a prior on the secret and an integer k and we are tasked with designing a channel which is optimal in the sense of leaking as little as possible.

[3]The fact that this is a worst-case assumption for the defender (and hence, provides stronger guarantees in terms of leakage) is closely related to results such as the *Gibbs*' inequality, or equivalently, the positivity of the *Kullback–Leibler* divergence.

[4]Adversaries can indeed learn the distribution of the secrets by observing a long enough history of the defender even for an i.i.d. source.

[5]Note that we take the cloaks to be "sets", hence, the implicit assumption is that always a uniformly randomized "permutation" of the elements of a cloak is also performed, so that any information associated with the order of elements is eliminated.

[6]We use the notation $a +_p b$ to represent an outcome which is a with probability p, and b with probability $1 - p$.

Entropies: Let $\mathcal{H}[\boldsymbol{\theta}] := H(\boldsymbol{P})$ denote the entropy of the random variable $\boldsymbol{\theta}$ with probability distribution $\boldsymbol{P}$, where the entropy function H is from the set of (discrete) probability distributions to the real numbers. In our setting, $\mathcal{H}[\boldsymbol{\theta}]$ is a measure of the *prior* uncertainty of the adversary about the secret. We consider a general entropy function H that only needs to satisfy a set of mild conditions as follows:

- **Symmetry.** $H(\boldsymbol{P})$ is invariant under permutations of $p_1, \ldots, p_n$. In other words, the entropy should depend only on the probability distribution of a random variable, and not on the specific labeling of them.
- **Expansibility.** Enlarging the secret space by adding zero probability secrets ("expansion" by zero components) should not change the entropy.
- **Core-Concavity.**[7] $H(\boldsymbol{P})$ is core-concave if it can be written as $H(\boldsymbol{P}) = \eta(F(\boldsymbol{P}))$, where $\eta : \mathbb{R} \to \mathbb{R}$ is a non-constant function on real numbers, F is a scalar function on probability distributions, and we have:[8]

$$\eta\text{: increasing, } F\text{: concave; or} \tag{2a}$$
$$\eta\text{: decreasing, } F\text{: convex.} \tag{2b}$$

Note that since η is just a non-constant univariate function, the symmetry and expansibility of H simply translate to F and vice versa. These conditions are not restrictive; in fact, the first two are among the axiomatic conditions of any entropy function [25]. The core-concavity property, as we defined above, is also a weak restriction and virtually all well-known entropy measures, some of which we discuss later, are indeed core-concave.

Next, we overview the concepts of *majorization* and *Schur-concavity*, which we use later in our analysis. First, some preliminaries: for a vector $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, let $\boldsymbol{a}^{\downarrow} = (a_{[1]}, \ldots, a_{[n]})$ denote a vector with the same elements but sorted in descending order, that is, $\boldsymbol{a}^{\downarrow}(i) = a_{[i]}$ represents the i'th-largest element of $\boldsymbol{a}$. Now, for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, we denote $\boldsymbol{a} \succ \boldsymbol{b}$ and say $\boldsymbol{a}$ *majorizes* $\boldsymbol{b}$ (or equivalently, $\boldsymbol{b}$ is *majorized* or *dominated* by $\boldsymbol{a}$) iff: $\sum_{i=1}^{j} a_{[i]} \geq \sum_{i=1}^{j} b_{[i]}$ for all $j = 1, \ldots, (n-1)$, and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *Schur-concave* iff: for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{a} \succ \boldsymbol{b}$ implies $f(\boldsymbol{a}) \leq f(\boldsymbol{b})$. From basic convex analysis, e.g. [26, Proposition 3.C.2], we have that every function that is symmetric and concave, is also Schur-concave. Therefore, the entropy functions H that we consider (and their corresponding F functions) are Schur-concave as well.

Some notable examples of frequently used entropies with practical interpretations are listed below. Note that all of these example entropy functions satisfy the symmetry, expansibility, and core-concavity properties:

- **1-Guess-Error-Probability.** $H(\boldsymbol{P}) = 1 - p_{[1]}$, which is the probability of failure of an adversary that gets to make one (optimal) guess about the secret. This measure is closely related to the more well-known **Min-Entropy**, where $H(\boldsymbol{P}) = -\log p_{[1]}$.
- ***l*-Guess-Error-Probability.** A generalization of the 1-guess-error-probability, $H(\boldsymbol{P}) = 1 - \sum_{i=1}^{l} p_{[i]}$, which is the probability of failure of an adversary that gets to make l (optimal) guesses.
- **Guesswork.** $H(\boldsymbol{P}) = \sum_{i=1}^{n} i p_{[i]}$, which is the expected number of tries of an (optimally guessing) adversary with unlimited number of allowed guesses before (and including) the correct one.
- **(Gibbs)-Shannon.** $H(\boldsymbol{P}) = -\sum_{i=1}^{n} p_i \log(p_i)$.
- **Rényi Entropy** A parametric family of entropies (for parameter α where $\alpha \geq 0, \alpha \neq 1$), defined as follows: $H_\alpha(\boldsymbol{P}) = \frac{1}{1-\alpha} \log(\sum_{i=1}^{n} p_i^\alpha)$, or equivalently, $H_\alpha(\boldsymbol{P}) = \frac{\alpha}{1-\alpha} \log \|\boldsymbol{P}\|_\alpha$, where $\|\boldsymbol{P}\|_\alpha$ denotes the α-norm of $\boldsymbol{P}$. Shannon and Min-Entropy can be derived as special cases by letting $\alpha \to 1$ and $\alpha \to \infty$, respectively. Two other well-known members of this family are *Collision* entropy for $\alpha = 2$: $H_2(\boldsymbol{P}) = -\log \sum_{i=1}^{n} p_i^2$, and *Hartley* entropy, for $\alpha = 0$: $H_0(\boldsymbol{P}) = \log |\operatorname{supp}(\boldsymbol{P})| = \log(n)$.

It is straightforward to see the symmetry and expansibility of each of the above entropy measures. Regarding core-concavity, note that for $1/l$-Guess-Error-Probability, Guesswork and Shannon entropies, η can be simply taken as the identity function, $\eta(x) = x$, which is increasing. This is because for all of these entropies, H is itself concave. For Rényi family, we can either take $\eta(x) = \frac{-1}{\alpha-1} \log(x)$ for $x > 0$ and $F(\boldsymbol{P}) = \|\boldsymbol{P}\|_\alpha^\alpha = \sum_{i=1}^{n} p_i^\alpha$, or $\eta(x) = \frac{-\alpha}{\alpha-1} \log(x)$ and $F(\boldsymbol{P}) = \|\boldsymbol{P}\|_\alpha$. Now, for both representations, if $\alpha \in [0, 1)$, η is increasing and F is concave, and if $\alpha > 1$, η is decreasing and F is convex. Hence, in both cases, the H function is core-concave.

The *posterior (conditional) entropy* of the secret, sometimes also referred to as the *equivocation*, is denoted by $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$, where $\boldsymbol{M}$ is the random variable associated with the observed outputs, here, cloaks. $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$ should be a measure of the uncertainty of the adversary about the secret on average *after* observing the cloaks. The reduction in the uncertainty of the adversary about the secret after observing the cloak, i.e., $\mathcal{H}[\boldsymbol{\theta}] - \mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$, is the leakage of information (see, e.g. [6]). Note that in our setting, the defender cannot choose or change the distribution of her secret, and therefore, irrespective of the choice of the entropy function, the prior entropy of the secret is unaffected by her cloaking strategy. Hence, the problem of minimizing information leakage becomes equivalent to maximizing the posterior entropy of the secret, i.e., solving: $\max_{\boldsymbol{\delta} \in \mathscr{D}} \mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$.

Consider an arbitrary strategy of the defender $\boldsymbol{\delta}$, and let $\mathcal{M}^+(\boldsymbol{\delta})$ be the set of all cloaks that each has a nonzero probability of being observed by the adversary, i.e., $\mathcal{M}^+(\boldsymbol{\delta}) = \cup_{\theta \in \Theta} \operatorname{supp}(\boldsymbol{\delta}(\theta))$. These are labeled as the "on-path" cloaks. We will omit the argument $\boldsymbol{\delta}$ whenever not

[7] No other potential meaning of the term "core-concavity" is intended.

[8] By definition, $F(\boldsymbol{P})$ is concave (resp. convex) in $\boldsymbol{P}$ iff for any $\lambda \in (0, 1)$ and $\boldsymbol{P}_1, \boldsymbol{P}_2 \in \Delta\Theta$, we have: $\lambda F(\boldsymbol{P}_1) + (1-\lambda) F(\boldsymbol{P}_2) \leq$ (resp. $\geq$) $F(\lambda \boldsymbol{P}_1 + (1-\lambda)\boldsymbol{P}_2)$.

ambiguous.

We assume that the conditional entropy for an entropy measure $H(\boldsymbol{P}) = \eta(F(\boldsymbol{P}))$ has the following structure:[9]

$$\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = \eta\Big(\sum_{M\in\mathcal{M}^+} \Pr(M) F\left(\boldsymbol{P}(\boldsymbol{\theta}|M)\right)\Big) \tag{3}$$

where $\Pr(M)$, short for $\Pr(\boldsymbol{M} = M)$, is the probability that cloak M is observed by the adversary, and $\boldsymbol{P}(\boldsymbol{\theta}|M)$ is the "posterior" distribution of the secrets given cloak M is observed, which is given by applying the *Bayes' rule*. Specifically, $\boldsymbol{P}(\boldsymbol{\theta} = \theta|M) := \Pr(\theta, M)/\Pr(M) = \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\Pr(M)$ where $\Pr(M) = \sum_{\theta'\in\Theta} \boldsymbol{P}(\theta')\boldsymbol{\delta}(M;\theta')$. Notice that we did not impose positivity of the entropy function H, as our results indeed do not rely on that. Moreover, our assumed properties are sufficient to establish that the leakage, defined as $\mathcal{H}[\boldsymbol{\theta}] - \mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$ is always positive. Specifically, suppose that case (2a) holds. Then, taking $\Pr(M)$ to be the coefficients of a convex combination (as they are positive and add up to one), the concavity of F directly gives:[10]

$$\sum_{M\in\mathcal{M}^+} \Pr(M) F\left(\boldsymbol{P}(\boldsymbol{\theta}|M)\right) \le F\left(\sum_{M\in\mathcal{M}^+} \Pr(M)\boldsymbol{P}(\boldsymbol{\theta}|M)\right)$$

The right hand side can be simplified to $F(\sum_{M\in\mathcal{M}^+} \boldsymbol{P}(\boldsymbol{\theta}, M)) = F(\boldsymbol{P})$. Moreover, since η is just an increasing univariate function, we have $\eta\left(\sum_{M\in\mathcal{M}^+} \Pr(M) F\left(\boldsymbol{P}(\boldsymbol{\theta}|M)\right)\right) \le \eta\left(F(\boldsymbol{P})\right)$, which is exactly the positivity of leakage. An almost identical argument applies for the (2b) cases.

Recall that for $1/l$-Guess-Error Probability, Guesswork and Shannon entropy, η was just the identity function. Hence, for these entropies, (3) simply reduces to:

$$\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = \sum_{M\in\mathcal{M}^+} \Pr(M) H(\boldsymbol{P}(\boldsymbol{\theta}|M)) \tag{4}$$

For Shannon entropy, the expression in (4) is exactly the classic conditional entropy. However, there is no commonly accepted definition of conditional entropy for other entropies. Nevertheless, the problem of maximizing the posterior entropy as defined in (4) have intuitive interpretations for a number of entropy measures that we introduced. Specifically, for the "1-Guess-Error-Probability", we show in Section V, this equivalently models a two-player zero-sum game of incomplete information in which the defender faces a strategic adversary with knowledge of the prior who can make at most one guess after observing the cloak, gain one unit if his guess is correct and zero otherwise. Similarly, the problem of maximizing the posterior entropy with respect to "l-Guess-Error-Probability" matches the problem of a defender in another two-player zero-sum game in which the adversary gets to make up to l guesses after observing the cloak and wins one unit if any of the guesses are correct and zero otherwise. In the same spirit, the corresponding two-player zero-sum game for the "Guesswork" entropy is one in which the adversary incurs a cost proportional to the number of guesses he takes before a correct guess. In Section V, besides establishing these connections to zero-sum games, we also show an interesting result: we provide a sufficient condition under which there is no pure strategy Nash Equilibrium of the corresponding games.

For Rényi entropies, the definition of conditional entropy as in (4), except for the limit case of $\alpha \to 1$, violates two desirable properties of "monotonicity" and "chain rule". Two commonly used alternative definitions that accommodate a set of desirable properties (ref. [27], [28]) are the following:

$$\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = \frac{-1}{\alpha-1}\log\Big(\sum_{M\in\mathcal{M}^+} \Pr(M)\|\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\alpha^\alpha\Big); \tag{5a}$$

$$\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = -\frac{\alpha}{\alpha-1}\log\Big(\sum_{M\in\mathcal{M}^+} \Pr(M)\|\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\alpha\Big). \tag{5b}$$

where, as before, $\|\cdot\|_\alpha$ denotes the α-norm. In particular, both of these definitions for $\alpha \to \infty$ yield the same definition for the conditional Min-Entropy: $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = -\log\left(\sum_{M\in\mathcal{M}^+} \Pr(M)\max(\boldsymbol{P}(\boldsymbol{\theta}|M)\right)$. Both variants of conditional entropy for Rényi family comply with the general structure of (3). In particular, for the first expression for conditional Rényi entropy (5a), $\eta(x) = \frac{-1}{\alpha-1}\log(x)$ for $x > 0$ and $F(\boldsymbol{P}) = \|\boldsymbol{P}\|_\alpha^\alpha = \sum_{i=1}^n p_i^\alpha$. For the alternative definition in (5b), we have $\eta(x) = \frac{-\alpha}{\alpha-1}\log(x)$ and $F(\boldsymbol{P}) = \|\boldsymbol{P}\|_\alpha$. Once again, recall that for $\alpha \in [0,1)$, for both cases η is increasing and F is concave, and for $\alpha > 1$, η is decreasing and F is convex.

As each entropy measure has its own distinct form and interpretation, it could have been the case that optimality of a cloaking scheme sensitively depend on the measure of entropy considered. However, in the next section, given $\boldsymbol{P}$ and k, we construct a cloaking strategy that is optimal with respect to any entropy measure that satisfies our four mild assumptions, and hence, in that sense, is *universally optimal*.

Table I
LIST OF THE MAIN NOTATIONS.

Parameter	Definition
Θ, n	Set of possible secrets, number of possible secrets: $n = \lvert\Theta\rvert \ge 3$.
$\boldsymbol{d}, \boldsymbol{\delta}$	A deterministic plan of action (pure strategy), and a randomized strategy of defender.
$\mathcal{M}(\theta)$	Set of permissible cloaks M for secret θ.
$\mathcal{M}$	Set of all permissible cloaks: $\cup_{\theta\in\Theta}\mathcal{M}(\theta)$.
k	The maximum allowed size of the cloaks, where $1 \le k < n$.

[9]Note that this structure for the conditional entropy is consistent with the unconditional entropy, since conditioning on an independent random variable gives $\mathcal{H}[\boldsymbol{\theta}] = H(\boldsymbol{P}) = \eta\left(F(\boldsymbol{P})\right)$. If η is strictly monotonic, (3) can also be written in terms of the unconditional entropy function, H, as: $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = \eta\left(\sum_{M\in\mathcal{M}^+} \Pr(M)\eta^{-1}\left(H\left(\boldsymbol{P}(\boldsymbol{\theta}|M)\right)\right)\right)$.

[10]Or equivalently, using the Jensen's inequality.

III. Analysis

In this section, we derive (in closed form) the maximum possible posterior entropy that can be achieved among all feasible randomized cloaking strategies for a given prior on the secrets $\boldsymbol{P}$, a cloak size cap k, and a measure of entropy H (Theorem 1-A). Our result is "constructive", in that, in Algorithm 1, we explicitly provide a cloaking strategy that achieves this maximum posterior entropy (and hence, minimum leakage) for any symmetric, expansible, core-concave measure of entropy (Theorem 1-B).

Before we present our formal result, let us get a feeling about the behavior of an optimal cloaking. Intuitively, the randomized strategy should try to induce posterior distributions over the secrets that are as close to uniform distribution as possible, since any well-defined measure of uncertainty increases as the distributions gets closer to uniform. The ideal case is that given any shown cloak, after the Bayesian update, the secret be equally likely any of the k members of the cloak, i.e., inducing uniform posteriors over the k elements of the cloak. However, if the prior distribution is too "skewed" and the cap size of the cloaks is small, then this might not be feasible, as the secrets with too big prior probabilities will still have higher posteriors. If a prior probability of a secret is too big to be made uniform in the posterior, i.e., a "giant", then it should be instead maximally leveraged against to hide other secrets in its "shadow". So, intuitively, an optimal strategy should try to induce posteriors that are uniform over as many of the "small" probability secrets as possible and the "giants" should always be included in the cloaks to provide "coverage" for the small probability secrets.

In order to formally present our results, we need to introduce some auxiliary parameters. Given k and $\boldsymbol{P} = (p_1, \ldots, p_n)$, sorted in descending order, let index J be defined as follows:

$$J := \min \left\{ j : 1 \leq j \leq k, p_j \leq \frac{\sum_{i=j}^{n} p_i}{k-j+1} \right\}. \qquad (6)$$

Note that for $j = k$, the condition $p_j \leq \sum_{j=1}^{n} p_i/(k-j+1)$ reduces to $p_k \leq \sum_{i=k}^{n} p_i$, which is trivially satisfied. Therefore, J is well-defined (i.e., can always be found), and we have $1 \leq J \leq k$. Along the lines of the above intuitive discussion, the first $J-1$ probabilities are the "elephant" secrets. Next, for a $j \in \{1, \ldots, k\}$ and prior distribution $\boldsymbol{P} = (p_1, \ldots, p_n)$, let $\boldsymbol{\pi_j}$ denote the probability distribution over k elements as the following:

$$\boldsymbol{\pi_j} := \left(p_1, \ldots, p_{j-1}, \frac{\sum_{i=j}^{n} p_i}{k-j+1}, \ldots, \frac{\sum_{i=j}^{n} p_i}{k-j+1}\right), \text{ i.e.:}$$

$$\boldsymbol{\pi_j}(l) = p_l : \; l \leq j-1, \quad \boldsymbol{\pi_j}(l) = \frac{\sum_{i=j}^{n} p_i}{k-j+1} : \; j \leq l \leq k \qquad (7)$$

In words, $\boldsymbol{\pi_j}$ is a k-sized probability distribution that is constructed by keeping the top $j-1$ probabilities of the prior as is, and then "wrapping" or "mashing" the remaining probabilities of the prior together and spreading them uniformly over the remaining $k-(j-1)$ elements. Note that $\boldsymbol{\pi_1}$ is simply the uniform distribution over the entire k elements. Finally, recall that we used $\mathcal{M}(\theta)$ to denote the set of feasible cloaks for secret θ, which is composed of all the subsets of size *at most* k that include θ. Now, we introduce the notation $\mathcal{M}^*(S)$, where $S \subseteq \Theta$ and $|S| \leq k$, to denote the set of feasible cloaks that include all the elements of S and have size *exactly equal to* k, i.e., are maximally sized. Formally, $\mathcal{M}^*(S) := \{M \subset \Theta : S \subseteq M, |M| = k\}$.[11] This notation is used in Step-2 of the Algorithm as well as in our proofs. For a simple example, suppose $\Theta = \{1,2,3,4\}$ and $k = 3$, then $\mathcal{M}^*(\{1\}) = \{\{1,2,3\}, \{1,2,4\}, \{1,3,4\}\}$, and $\mathcal{M}^*(\{1,2\}) = \{\{1,2,3\}, \{1,2,4\}\}$, and so on.

We are now ready to express our main result:

Theorem 1: Let $\boldsymbol{P} = (p_1, \ldots, p_n)$ be the prior (sorted in descending order), and let k be the maximum permissible size of the cloaks. Let index J and probability distributions $\boldsymbol{\pi_j}$ be defined as in (6) and (7), respectively. Let the entropy be measured by the symmetric, expansible and core-concave function H. Then:

A. The maximum achievable posterior entropy among all (potentially randomized) cloaking strategies is $H(\boldsymbol{\pi_J})$.
B. Algorithm 1 explicitly provides a feasible (randomized) cloaking strategy that achieves the above maximum posterior entropy for any choice of the entropy, and hence, in this sense, is universally optimal.[12]

Algorithm 1: Optimal Cloaking Strategy for a given $\boldsymbol{P}$, k (Theorem 1)

Input: $\boldsymbol{P} = (p_1, \ldots, p_n)$ in descending order, k
Output: $\boldsymbol{\delta}(M;\theta)$ for $\forall \theta \in \Theta, \forall M \in \mathcal{M}$

1: **Find** $J \leftarrow \min \left\{ 1 \leq j \leq k : p_j \leq \frac{\sum_{i=j}^{n} p_i}{k-j+1} \right\}$

2: **Solve** $\sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M = p_i, \quad \forall i = J, \ldots, n$
s. t.: $x_M \geq 0, \quad \forall M \in \mathcal{M}^*(\{1, \ldots, J-1\})$

3: $\boldsymbol{\delta}(M;i) \leftarrow x_M/p_i$ $\quad \forall i = J, \ldots, n$, $\forall M \in \mathcal{M}^*(\{1, \ldots, J-1, i\})$

4: $\boldsymbol{\delta}(M;i) \leftarrow x_M(k-J+1)/\sum_{j=J}^{n} p_j$ $\quad \forall i = 1, \ldots, J-1$, $\forall M \in \mathcal{M}^*(\{1, \ldots, J-1\})$

5: $\boldsymbol{\delta}(M;\theta) \leftarrow 0$ $\quad$ *everywhere else*

Although the Algorithm may look cryptic, it is essentially

[11] Note that of course both $\mathcal{M}(\theta)$ and $\mathcal{M}^*(S)$ depend on the parameter k, however, since k does not vary in our usage of these two notations and this dependence is clear from the context, we have not made it explicit.

[12] Note that the optimal strategy may not be unique, since the set of positive solutions to the linear system of equations in Step 2 of Algorithm 1 are in general convex polyhedra. The theorem guarantees that all of such solutions are universally optimal.

doing something simple: it generates randomized strategies $\boldsymbol{\delta}(M;\theta)$ such that given any cloak M that is shown to the adversary, the posterior distribution over the secrets in the cloak is exactly $\boldsymbol{\pi_J}$ (following a straightforward use of Bayes' rule). It does so by *always* including the $\theta_1,\ldots,\theta_{J-1}$ in the cloak, and carefully choosing the randomization of the cloaks such that the posterior probability over the rest of the remaining $k-J+1$ items is uniform (guaranteed by the solution of the linear system of equations in Step 2), and the posterior distribution over the first $J-1$ secrets is exactly their priors (guaranteed by steps 3 and 4). The proof of the theorem is provided in Appendix A in its entirety. It follows simple logical steps but is nevertheless non-trivial. In particular, first, we establish that $H(\boldsymbol{\pi_J})$ is an upper-bound for the posterior Entropy, that is, no feasible cloaking scheme can increase the posterior uncertainty of the adversary above $H(\boldsymbol{\pi_J})$. Subsequently, we show that Algorithm 1 provides a *feasible* cloaking strategy that *achieves this upper-bounds*, and hence is optimal. A tricky step of the proof is to show that the linear system of equations in Step 2 indeed has a positive solution.

Here, we compute the optimal strategies through a few toy examples to gain some intuition about the algorithm. Consider the case of four possible secrets $1,2,3,4$ and $k=3$. We have then the following possible size 3 cloaks:

$$M_1=\{1,2,3\}, M_2=\{1,2,4\}, M_3=\{1,3,4\}, M_4=\{2,3,4\}$$

Consider the following three possible priors over the secrets: $\boldsymbol{P}_1 = (0.3, 0.28, 0.22, 0.2)$, $\boldsymbol{P}_2 = (0.36, 0.3, 0.2, 0.14)$, $\boldsymbol{P}_3 = (0.4, 0.35, 0.15, 0.1)$. For $\boldsymbol{P}_1$, we have: $\boldsymbol{P}_1(1) = 0.3 \leq 1/k = 1/3 = 0.33$, hence $J=1$, and the optimal strategy will induce $(1/3, 1/3, 1/3)$ posterior distributions. For $\boldsymbol{P}_2$, we have: $\boldsymbol{P}_2(1) = 0.36 > 1/k$ but $\boldsymbol{P}_2(2) = 0.3 \leq (0.3+0.2+0.14)/(k-1) = 0.64/2 = 0.32$, therefore $J=2$, and the optimal strategy will always include 1 in the cloak and induce $(0.36, 0.32, 0.32)$ posterior distributions. For $\boldsymbol{P}_3$: $\boldsymbol{P}_3(1) = 0.4 > 1/k$, $\boldsymbol{P}_3(2) = 0.35 > (0.35+0.15+0.1)/(k-1) = 0.6/2 = 0.3$, and only $\boldsymbol{P}_3(3) = 0.15 \leq (0.15+0.1)/(k-2) = 0.25/1 = 0.25$, thus $J=3$, and the optimal strategy will always include 1 and 2 in the cloak and induce $(0.4, 0.35, 0.25)$ posterior distributions. Then the corresponding optimal cloaking strategies for $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3$ respectively, as given by Algorithm 1 are the following:

	1 0.3	2 0.28	3 0.22	4 0.2
M_1:	0.4444	0.4762	0.6061	0
M_2:	0.3778	0.4048	0	0.5667
M_3:	0.1778	0	0.2424	0.2667
M_4:	0	0.1190	0.1515	0.1667

	1 0.36	2 0.3	3 0.2	4 0.14
M_1	0.5625	0.6000	0.9000	0
M_2	0.3750	0.4000	0	0.8571
M_3	0.0625	0	0.1000	0.1429
M_4	0	0	0	0

	1 0.4	2 0.35	3 0.15	4 0.1
M_1	0.6000	0.6000	1.0000	0
M_2	0.4000	0.4000	0	1.0000
M_3	0	0	0	0
M_4	0	0	0	0

Recall from Bayes' rule that $\boldsymbol{P}(\boldsymbol{\theta} = \theta|M) := \Pr(\theta, M)/\Pr(M) = \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\Pr(M)$. For instance, we can check that the optimal cloaking strategy for $\boldsymbol{P}_1$, irrespective of the shown cloak, induces uniform distribution over its 3 elements. Since the denominator is the same, we just need to verify $\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)$ is the same for all $\theta \in M$. For instance, for $M_1=\{1,2,3\}$ we have: $0.3\times 0.4444 = 0.28\times 0.4762 = 0.22\times 0.6061 = 0.1333$. Similarly, for $M_2=\{1,2,4\}$ we have: $0.3\times 0.3778 = 0.28\times 0.4048 = 0.2\times 0.5667 = 0.1133$. And finally, for $M_3=\{1,3,4\}$, we have: $0.3\times 0.1778 = 0.22\times 0.2424 = 0.2\times 0.2667 = 0.0533$.

Discussion: complexity of computing a solution: Referring to Algorithm 1, if the prior probabilities are already sorted, then the complexity of finding an optimal cloaking strategy is determined by the complexity of finding J, which is upper-bounded by n, just by an exhaustive search. However, the following simple lemma shows that in fact, given a list of partial sums of $\boldsymbol{P}$, a binary search can be used, yielding an $O(\log(n))$ worst-case complexity.

Lemma 1: Suppose (as before) the prior $\boldsymbol{P} = (p_1,\ldots,p_n)$ is sorted in descending order. If $p_j \leq (>)\frac{\sum_{i=j}^{n} p_i}{k-j+1}$, then $p_m \leq (>)\frac{\sum_{i=m}^{n} p_i}{k-m+1}$ for all $m \leq (>) j$, (respectively).

IV. Departure From Symmetric Entropies: Extension to General Gain-Based Measures

In the previous section, we provided a probabilistic cloaking strategy that yields minimum leakage with respect to a large class of classical entropy measures. Our analysis only relied on structural properties of the entropy function, namely: symmetry, expansibility, and what we called "core-concavity". A major point of departure from this family of entropies, where potentially all three of these properties can be violated, is the g-entropy introduced in [9]. g-entropy is a generalization of the notion of Min-Entropy by allowing secret/guess dependent gains to a guessing adversary. This notion of leakage has received attraction in the security research community (e.g. [4], [10], [29]–[31]). In what follows, we first overview the notion of g-entropy and g-leakage, then introduce our generalization of g-entropy by fusing it with Rényi entropies, and present an extension of our main theorem.

Recall that in the case of Min-Entropy, the entropy is related to the expected gain of an adversary that makes one guess and wins one unit for every correct guess and zero otherwise, irrespective of the actual secret. Now consider instead that the adversary gains (and the defender loses) a reward of $g(w,\theta) \in [0,1]$ units if the guess of the adversary

and the actual secret are $w \in \mathcal{W}$ and $\theta \in \Theta$, respectively. We can make the "natural assumption" that, among all guesses, the highest gain is still achieved for a "correct" guess, but now the value of the gain for correct guess may depend on the secret. Moreover, the adversary may gain some value for "close" guesses, albeit less than the gain for the correct guess. The space of guesses $\mathcal{W}$ can be simply the same as the space of secrets Θ, or more complicated sets such as all permutations of the secrets for allowing a generalization of guesswork entropy through the gain function. The g-entropy is then defined as the negative of the logarithm of the maximum expected reward of the adversary. Specifically, if the random variable $\boldsymbol{\theta}$ has distribution $\boldsymbol{P}$, then the (prior) g-entropy of $\boldsymbol{\theta}$ is given as:

$$\mathcal{H}_g[\boldsymbol{\theta}] = H_g(\boldsymbol{P}) = -\log\left(\max_{w\in\mathcal{W}}\left(\sum_{\theta\in\Theta} g(w,\theta)\boldsymbol{P}(\theta)\right)\right)$$

Representing the gain function g in a matrix $G \in \mathbb{R}^{|\mathcal{W}|\times|\Theta|}$, where $G_{w,\theta} := g(w,\theta)$, we can rewrite H_g using the infinity-norm as follows:

$$H_g(\boldsymbol{P}) = -\log\|G\boldsymbol{P}\|_\infty$$

The posterior g-entropy after observing output random variable $\boldsymbol{M}$ is defined as (see [9]):

$$\mathcal{H}_g(\boldsymbol{\theta}|\boldsymbol{M}) = -\log\Big(\sum_{M\in\mathcal{M}^+}\Pr(M)\max_{w\in\mathcal{W}}\big(\sum_{\theta\in\Theta} g(w,\theta)\boldsymbol{P}(\theta|M)\big)\Big)$$

which, using the matrix representation G can be written as:

$$\mathcal{H}_g(\boldsymbol{\theta}|\boldsymbol{M}) = -\log\left(\sum_{M\in\mathcal{M}^+}\Pr(M)\|G\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\infty\right)$$

The difference between prior and posterior g-entropies is then defined as the *g-leakage*.

The matrix representation of the gain function makes it specially evident that the g-entropy is primarily an extension of the Min-Entropy, which corresponds to the case in which G is chosen to be the $|\Theta| \times |\Theta|$ identity matrix. However, as pointed out in [4], l-guess and guesswork entropies can also be recovered as special cases of g-entropy if the space of guesses $\mathcal{W}$ is respectively expanded to the set of all subsets of Θ with cardinality l, and the set of all $(n!)$ permutations of Θ, and the gain entries are selected appropriately. Specifically, for l-guess, $g(w,\theta) = 1$ if $\theta \in w$ and is zero otherwise. For guesswork entropy, the space of gain values should be extended to include negative numbers, specifically, $g(w,\theta) = -i$, where i is the index (i.e., the position) of θ in the permutation w. However, it is not possible to capture Shannon entropy as a g-entropy with a g function with countable range anymore.[13] Next, we introduce an extension of g-entropy that recaptures not only Shannon, but also all of the Rényi family as special cases.

Comparing the ∞-norm representation of the Min-Entropy: $-\log\|\boldsymbol{P}\|_\infty$ with that of the g-entropy: $-\log\|G\boldsymbol{P}\|_\infty$, and noting that Min-Entropy is a limit member of the Rényi entropies for $\alpha \to \infty$, we introduce an (α, g) family of entropies for $\alpha \geq 0$, $\alpha \neq 1$, as follows:

$$\mathcal{H}_{\alpha,g}(\boldsymbol{\theta}) = H_{\alpha,g}(\boldsymbol{P}) := H_\alpha(G\boldsymbol{P}) = \frac{\alpha}{1-\alpha}\log\|G\boldsymbol{P}\|_\alpha \quad (8)$$

where H_α is just the usual Rényi entropy for parameter α. Moreover the conditional (α, g) entropy $\mathcal{H}_{\alpha,g}[\boldsymbol{\theta}|\boldsymbol{M}]$, is also set to $\mathcal{H}_\alpha[\boldsymbol{\theta}|\boldsymbol{M}]$ according to one of the forms in (5a) or (5b) with $\boldsymbol{P}(\boldsymbol{\theta}|\boldsymbol{M})$ replaced by $G\boldsymbol{P}(\boldsymbol{\theta}|\boldsymbol{M})$. All Rényi entropies are trivially instances of (α, g) family by taking G to be the identity matrix. In particular, Shannon entropy is retrieved by also letting $\alpha \to 1$. The (α, g)-leakage can be defined as the difference between prior and posterior entropies. In particular, g-leakage defined in [9] is obtained from (α, g)-leakage by letting $\alpha \to \infty$.

More generally, for a symmetric, expansible and core-concave entropy function $H(\boldsymbol{P})$, it may be possible to define a corresponding gain-based entropy function as $H_g(\boldsymbol{P})$ for a given gain function g with a matrix representation G, by replacing $\boldsymbol{P}$ with $G\boldsymbol{P}$ in its formulation, i.e., by taking $H_g(\boldsymbol{P}) := H(G\boldsymbol{P})$, as long as $G\boldsymbol{P}$ is in the domain of the H function. We will then refer to H as the *underlying* symmetric entropy.[14] In particular, we consider a general family of entropies $\mathcal{H}[\boldsymbol{\theta}]$ that can be expressed as $H_g(\boldsymbol{P}) = H(G\boldsymbol{P}) = \eta(F(G\boldsymbol{P}))$, and its conditional entropy can be written as:

$$\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}] = \eta\left(\sum_{M\in\mathcal{M}^+}\Pr(M)F(G\boldsymbol{P}(\boldsymbol{\theta}|M))\right) \quad (9)$$

where functions η and F satisfy (2a) or (2b). For instance, by taking $\eta(x) = -\log(x)$ and $F(\boldsymbol{P}) = \|\boldsymbol{P}\|_\infty$, we get the g-entropy [9].

Recall that in our setting, since the defender is unable to affect the prior distribution of the secret, the problem of achieving minimum leakage becomes equivalent to maximizing the posterior entropy. Note that for almost any G other than the identity matrix, the new entropy function $H_g(\boldsymbol{P}) = H(G\boldsymbol{P}) = \eta(F(G\boldsymbol{P}))$ is no longer symmetric in $\boldsymbol{P}$, and hence our main result does not directly apply here. However, for an important class of matrix gains, namely diagonal matrices, we present a generalization of Theorem 1. Specifically, for a given prior distribution $\boldsymbol{P}$, diagonal gain matrix G, and the cloak size cap k, we provide a probabilistic cloaking strategy that is leakage-optimal with

[13] In [4], the authors cast Shannon entropy as g-leakage but only by extending the space of guesses to be the uncountably infinite set of all probability distributions over the secret space.

[14] Note that $G\boldsymbol{P}$ may no longer be a probability distribution, but as long as the transformation of the probability simplexes under G is in the domain of the H function, $H_g(\boldsymbol{P})$ will simply be the evaluation of the H function at $G\boldsymbol{P}$. Moreover, we remind the reader that our results do not rely on the sign of the H function. However, following similar steps as after (3), the "leakage" is nevertheless positive (cf. [9, Theorem 4.1]).

respect to any entropy whose conditional entropy follows the structure of (9). In particular, it is optimal for any choice of α in the (α, g)-family, and hence with respect to g-leakage as a special case.

We use the notation $G = \mathrm{diag}(\boldsymbol{\gamma})$ where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^{+n}$, to indicate that G is a square diagonal matrix (zero for every entry except for possibly the diagonal elements). This models cases where the adversary gains $\gamma_i \geq 0$ if the secret is θ_i and he identifies it correctly, and zero if he mis-identifies. Although investigating only diagonal gain matrices may appear restrictive, they do capture the secret-dependent non-symmetric essence of the g-leakage notion. In Section V, using a game theoretic approach, we describe an LP that yields optimal cloaking strategies for a general G (although only for $\alpha = \infty$, i.e., g-leakage).

Theorem 2: Let $\boldsymbol{P} = (p_1, \ldots, p_n)$ be the prior, $G = \mathrm{diag}(\boldsymbol{\gamma})$, be the diagonal gain matrix, and let k be the size cap of the cloaks. Without loss of generality, assume that $G\boldsymbol{P} = (\gamma_1 p_1, \ldots, \gamma_n p_n)$ is in descending order, i.e., $(G\boldsymbol{P})_{[i]} = \gamma_i p_i$. Let index J_g and vector $\boldsymbol{\pi_{g,j}} \in \mathbb{R}^{+k}$ be defined as follows:

$$J_g := \min \left\{ j : 1 \leq j \leq k, \gamma_j p_j \leq \frac{\sum_{i=j}^{n} \gamma_i p_i}{k - j + 1} \right\}$$

$$\boldsymbol{\pi_{g,j}}(l) := \gamma_l p_l : \; l \leq j-1, \; \boldsymbol{\pi_{g,j}}(l) = \frac{\sum_{i=j}^{n} \gamma_i p_i}{k - j + 1} : j \leq l \leq k$$

Consider an entropy function $H_g(\boldsymbol{P}) = H(G\boldsymbol{P}) = \eta(F(G\boldsymbol{P}))$ with conditional entropy as in (9) where functions η and F satisfy (2a) or (2b), which includes our (α, g)-entropies as defined in (8). Then:

A. The maximum achievable posterior entropy $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}]$ among all cloaking strategies is $H(\boldsymbol{\pi_{g,J_g}})$.

B. If sorted $G\boldsymbol{P}$ (in descending order) instead of sorted $\boldsymbol{P}$ is passed to Algorithm 1 as its input argument, it explicitly provides a feasible (randomized) cloaking strategy that achieves the above maximum posterior entropy for any leakage measure of this family, in particular, g-leakage.

The proof of the theorem is provided in Appendix B.

The extension makes intuitive sense: The gain coefficients, γ_is, are a measure of the "relative importance" of having a secret revealed. The algorithm multiplies each probability by its corresponding gain and tries to make this effective importance of the secrets as equal as possible. Also note that, of course, this theorem reduces to Theorem 1 when G is the identity matrix.

V. Game Theoretic Analysis: Extension to General Cloaking Constraints and Gain Matrices

In this section, we show how the designing of leakage-optimal channels with respect to g-entropies can be cast as 2-player zero-sum games (2PZSG in short) of incomplete information. This observation enables us to develop a Linear Program that gives the minimum leakage and corresponding optimal strategies for a general gain matrix G and any cloaking constraints, beyond just the size-cap case that we investigated so far. As we mentioned earlier, Min-Entropy, Guesswork and l-guess entropies can all be expressed as special cases of g-entropy, so this section's formulation generalizes our result for these entropies as well.

The setup should be now familiar: Consider an adversary that makes a guess $w \in \mathcal{W}$ after observing a cloak M. A deterministic plan of action for such an adversary, denoted by $\boldsymbol{a}$, is hence a function from $\mathcal{M}$ to $\mathcal{W}$, specifying his guess given each (permissible) cloak. Hence, the space of the adversary's pure strategies is $\mathcal{A} := \mathcal{W}^{\mathcal{M}}$, i.e., the set of all functions from $\mathcal{M}$ to $\mathcal{W}$.

Similar to a randomized cloaking strategy of the defender, a randomized strategy of the adversary, designated by $\boldsymbol{\alpha}$, assigns a potentially probabilistic guess to each cloak, i.e., $\boldsymbol{\alpha} : \mathcal{M} \to \Delta\mathcal{W}$. Hence the space of randomized strategies of the adversary is simply $\mathscr{A} := (\Delta\mathcal{W})^{\mathcal{M}}$, i.e., the set of all functions from permissible cloaks to the probability distributions over guess choices. A pure and randomized *strategy profile* of the game are the pairs $(\boldsymbol{d}, \boldsymbol{a}) \in (\mathcal{D} \times \mathcal{A})$ and $(\boldsymbol{\delta}, \boldsymbol{\alpha}) \in (\mathscr{D} \times \mathscr{A})$, respectively. Recall that we used the notation $\boldsymbol{\delta}(M; \theta)$ to designate the probability at which, under the randomized strategy of $\boldsymbol{\delta}$, the defender chooses cloak M when her secret is θ. Likewise, $\boldsymbol{\alpha}(w; M)$ denotes the probability at which the adversary makes his guess to be w after observing M.

The outcome of each instance of the game is the following: the adversary wins (and the defender loses) $g(w, \theta)$ if adversary's guess and the (realization) of the secret had been w and θ respectively. The payoff of the game can in general be represented by the function $v : \Theta \times \mathcal{M} \times \mathcal{W} \to \mathbb{R}$. In our game, in particular, we have $v(\theta, M, w) = g(w, \theta)$. The payoff of the defender is the negative of the adversary's.

Recall that we used $\mathcal{M}(\theta)$ to denote the set of feasible cloaks that the defender can select for secret θ. In the previous section, we considered an important case of such constraints: that the cloak can be any subset of the secrets that include θ, as long as the size of the cloak does not exceed a given cap k. In our game formulation, we allow these cloaking constraints to be arbitrary, and in the most general form, given explicitly as $\mathfrak{M} = \mathcal{M}(\theta_1) \times \ldots \times \mathcal{M}(\theta_n)$.

Let $V(\boldsymbol{P}, \mathfrak{M}, \boldsymbol{\delta}, \boldsymbol{\alpha})$, which we show by V for brevity, represent the expected payoff of the adversary (to be maximized by him, and minimized by the defender). The expectation is taken with respect to the random realization of the secret according to $\boldsymbol{P}$ as well as any randomization present in the strategies of the two players.[15] Hence, in our problem:

$$V = \sum_{\theta \in \Theta} \sum_{M \in \mathcal{M}} \sum_{w \in \mathcal{W}} \boldsymbol{P}(\theta) \boldsymbol{\delta}(M; \theta) \boldsymbol{\alpha}(w; M) g(w, \theta) \quad (10)$$

[15]Note that pure strategies can be emulated as special cases of randomized strategies (where the distribution is degenerate).

As before, let $\mathcal{M}^+$ be the set of cloaks that (under the defender's strategy $\boldsymbol{\delta}$) each has a strictly positive probability of being observed by the adversary, i.e., $\mathcal{M}^+ = \cup_{\theta\in\Theta}\,\mathrm{supp}(\boldsymbol{\delta}(\theta))$. Since only these "on-path" cloaks contribute to the expected utilities, we have:

$$V = \sum_{\theta\in\Theta}\sum_{M\in\mathcal{M}^+}\sum_{w\in\mathcal{W}} \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)\boldsymbol{\alpha}(w;M)g(w,\theta)$$

$$= \sum_{M\in\mathcal{M}^+} \boldsymbol{P}_{\boldsymbol{\delta}}(M)\sum_{w\in\mathcal{W}}\boldsymbol{\alpha}(w;M)\sum_{\theta\in M} g(w,\theta)\frac{\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)}{\boldsymbol{P}_{\boldsymbol{\delta}}(M)}$$

where: $\boldsymbol{P}_{\boldsymbol{\delta}}(M) := \sum_{\theta\in\Theta}\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)$, that is the probability that M is observed by the adversary (which is nonzero for $M\in\mathcal{M}^+$). Note that $\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\boldsymbol{P}_{\boldsymbol{\delta}}(M)$ is the posterior probability that the secret is θ given that the observed cloak is M where the update in belief is done using the Bayes' rule. Let us denote the Bayesian update of the distribution of the secret given the observation of M by $\boldsymbol{P}^B(\cdot|M)$, that is, $\boldsymbol{P}^B(\theta|M) := \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\boldsymbol{P}_{\boldsymbol{\delta}}(M)$. Therefore, the expression for the expected payoff of the adversary can be written as:

$$V = \sum_{M\in\mathcal{M}^+} \boldsymbol{P}_{\boldsymbol{\delta}}(M)\sum_{w\in\mathcal{W}}\boldsymbol{\alpha}(w;M)\sum_{\theta\in\Theta} g(w,\theta)\boldsymbol{P}^B(\theta|M)$$

Given a strategy of the defender $\boldsymbol{\delta}$, let the highest value of V achieved when the adversary adopts a best response strategy to $\boldsymbol{\delta}$ be denoted by $\overline{V}$, which depends on $\boldsymbol{P}$, $\mathfrak{M}$ and $\boldsymbol{\delta}$. From the above expression for V, any best response strategy of the adversary must select a maximizer of the conditional expectation of the gain given each "on-path" cloak, that is:

$$\overline{V}(\boldsymbol{P},\mathfrak{M},\boldsymbol{\delta}) = \sum_{M\in\mathcal{M}^+}\boldsymbol{P}_{\boldsymbol{\delta}}(M)\max_{w\in\mathcal{W}}\left\{\sum_{\theta\in\Theta} g(w,\theta)\boldsymbol{P}^B(\theta|M)\right\}$$

Using the matrix representation of the gain function and the notion of infinity norm, this can be written simply as $\sum_{M\in\mathcal{M}^+}\boldsymbol{P}_{\boldsymbol{\delta}}(M)\|G\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\infty$. The value of $\overline{V}(\boldsymbol{P},\mathfrak{M},\boldsymbol{\delta})$ quantifies the worst expected loss of the defender given her randomized cloaking strategy $\boldsymbol{\delta}$. Let $\overline{V}^*(\boldsymbol{P},\mathfrak{M})$ denote its minimum over all feasible randomized cloaking strategies:

$$\overline{V}^*(\boldsymbol{P},\mathfrak{M}) = \min_{\boldsymbol{\delta}\in\mathscr{D}}\sum_{M\in\mathcal{M}^+}\boldsymbol{P}_{\boldsymbol{\delta}}(M)\|G\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\infty \qquad (11)$$

This is the *minimax* problem of the defender.

Recall the notion of g-entropy: $\mathcal{H}_g[\boldsymbol{\theta}] = -\log\|G\boldsymbol{P}\|_\infty$ and $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}] = -\log\left(\sum_{M\in\mathcal{M}^+}\boldsymbol{P}_{\boldsymbol{\delta}}(M)\|G\boldsymbol{P}(\boldsymbol{\theta}|M)\|_\infty\right)$. The connection should now be clear, that: finding the optimal (randomized) cloaking strategy that yields the least information leakage through cloaking (subject to cloaking constraints) with respect to g-leakage is equivalent to the minimax problem of the defender as stated in (11), where the defender, in a 2PZSG, faces a strategic adversary with knowledge of the prior that makes guesses about the secret after observing the cloak who gains $g(w,\theta)$ units if his guess is w and the actual secret is θ.

In any two-person game, a strategy pair is a (Nash) equilibrium if none of the players have any strictly advantageous unilateral deviation. In other words, keeping the strategy of the other player fixed, the strategy of each player must be a maximizer of its expected utility. Our first result shows that for an important class of cloaking constraints, deterministic strategy profiles never constitute a (Nash) equilibrium (except when the whole secret space is a permissible cloak for all secrets), and hence, any equilibrium (and thus, any optimal channel design as well as adversaries guessing strategy) must involve randomization. Suppose the cloaking constraints are such that $\forall\theta_i,\theta_j\in\Theta$, we have: $\mathcal{M}(\theta_i)\cap\mathcal{M}(\theta_j)\neq\emptyset$. In words, any two secrets have at least one permissible cloak in common. We refer to this property as *permissible connectivity*. Note that the "size-capped" cloak scenarios satisfy this property for any $k\geq 2$.

Proposition 3: If the cloaking constraints satisfy "permissible connectivity" and the gain matrix is diagonal with positive entries, then except for the trivial case of $\Theta\in\mathcal{M}(\theta)$ $\forall\theta$, there is no equilibrium solution among pure strategy profiles.

The proof of the proposition is provided in Appendix C. Intuitively, in such problems, even the "best" non-probabilistic strategies of the defender can be strictly improved (the uncertainty of the adversary strictly increased) by injecting ambiguity through randomization into it, as there is "room" for such maneuvers and the adversary cannot "corner" the defender, thanks to the "permissible connectivity".

The above result shows that the search for equilibria must be extended to randomized strategies. Existence of a solution among randomized strategy profiles is guaranteed by (an extension of) Nash's Theorem (or the duality in LPs).

Referring back to (11), since for $M\in\mathcal{M}^+$ we have $P_{\boldsymbol{\delta}}(M)\boldsymbol{P}^B(\theta|M) = \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)$, the minimax problem of the defender can be simply re-written as:

$$\overline{V}^*(\boldsymbol{P},\mathfrak{M}) = \min_{\boldsymbol{\delta}\in\mathscr{D}}\sum_{M\in\mathcal{M}}\max_{w\in\mathcal{W}}\left\{\sum_{\theta\in\Theta} g(w,\theta)\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)\right\}$$

Consider a defender strategy $\boldsymbol{\delta}^*$ that achieves the above optimization, that is, $\boldsymbol{\delta}^*\in\arg\max_{\boldsymbol{\delta}\in\mathscr{D}}\overline{V}(\boldsymbol{P},\mathfrak{M},\boldsymbol{\delta})$. Then by adopting strategy $\boldsymbol{\delta}^*$, she can "guarantee" that her expected cost will never be above $\overline{V}^*(\boldsymbol{P},\mathfrak{M})$ irrespective of the strategy of the adversary. Introducing auxiliary variables $\boldsymbol{v} = (v_M)$ for $M\in\mathcal{M}$, and $\boldsymbol{x} = (x_M^\theta) := \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)$ for all $M\in\mathcal{M}(\theta)$ and $\theta\in\Theta$ (or equivalently, for all $\theta\in M$ and $M\in\mathcal{M}$), the above minimax optimization problem can be cast as a linear program (LP):

$$\overline{V}^*(\boldsymbol{P},\mathfrak{M}) = \min_{\boldsymbol{x},\boldsymbol{v}}\sum_{M\in\mathcal{M}} v_M$$

$$\text{s.t.:}\quad v_M \geq \sum_{\theta\in\Theta} g(w,\theta)x_M^\theta,\quad \forall w\in\mathcal{W},\ \forall M\in\mathcal{M}$$

$$x_M^\theta\geq 0, \forall M\in\mathcal{M}(\theta),\forall\theta\in\Theta,\quad \sum_{M\in\mathcal{M}(\theta)} x_M^\theta = \boldsymbol{P}(\theta),\ \forall\theta\in\Theta$$

Specifically, a solution of the above LP denoted by $\boldsymbol{x}^* = ({x^*}^{\theta}_{M})$ provides the randomized equilibrium strategy of the defender through the transformation: $\boldsymbol{\delta}^*(M;\theta) = {x^*}^{\theta}_{M}/\boldsymbol{P}(\theta)$ for $M \in \mathcal{M}(\theta)$, $\theta \in \Theta$, and $\boldsymbol{\delta}^*(M;\theta) = 0$ for $M \notin \mathcal{M}$, $\theta \in \Theta$. Introducing variables $\boldsymbol{u} = (u_\theta)$ for $\theta \in \Theta$ and $\boldsymbol{y} = (y_w^M)$ for $w \in \mathcal{W}$, $M \in \mathcal{M}$, the dual of the above LP is:

$$\underline{V}^*(\boldsymbol{P}, \mathfrak{M}) := \max_{\boldsymbol{y},\boldsymbol{u}} \sum_{\theta\in\Theta} \boldsymbol{P}(\theta) u_\theta$$

$$\text{s.t.:} \quad u_\theta \le \sum_{w\in\mathcal{W}} g(w,\theta) y_w^M, \ \forall M \in \mathcal{M}(\theta), \forall \theta \in \Theta$$

$$y_w^M \ge 0, \forall w \in \mathcal{W}, \forall M \in \mathcal{M}, \quad \sum_{w\in\mathcal{W}} y_w^M = 1, \ \forall M \in \mathcal{M}$$

Let $\boldsymbol{y}^* = ({y^*}_w^M)$ be the solution of the dual LP (above). By adopting strategy $\boldsymbol{\alpha}^*(w, M) = {y^*}_w^M$ for all $M \in \mathcal{M}$ and $w \in \mathcal{W}$, the adversary can "guarantee" that his expected reward will never fall below $\underline{V}^*(\boldsymbol{P}, \mathfrak{M})$ irrespective of the cloaking strategy of the defender. Note that from strong duality of LPs, we have $\overline{V}^*(\boldsymbol{P}, \mathfrak{M}) = \underline{V}^*(\boldsymbol{P}, \mathfrak{M})$ which we denote by $V^*(\boldsymbol{P}, \mathfrak{M})$. This in part implies that $(\boldsymbol{\delta}^*, \boldsymbol{\alpha}^*)$ such found is indeed a randomized (Nash) equilibrium.

The above primal-dual LPs provides mutually optimal cloaking strategies of the defender and identification strategies of the adversary for any general cloaking constraints and G-matrix for the 2PZSG corresponding to g-leakage.

When the cloaking constraints are size-capped (to size at most k), any of the strategies given in Theorem 2 is also a NE and a minimax strategy of the defender. Our game-theoretic formulation provides the optimal cloaking strategies for a general (non-diagonal) gain matrix in g-leakage and arbitrary cloaking constraints that include the size-capped model of Theorem 2. However, it requires solving a linear optimization as opposed to the much simpler problem of finding any positive solution for the linear system of equations in Theorem 2 (with less number of variables). Moreover, we showed that our strategies in Theorems 1 and 2 are optimal with respect to any measure of entropy that satisfy mild conditions. In contrast, the game-theoretic approach requires solving a new LP per each entropy. For instance, for "Guesswork" entropy, the action space of the adversary (his feasible guesses) is the set of all the permutations of secrets, and hence, the size of the linear programming in our game theoretic model grows exponentially with the size of the secret space. Finally, unlike the LP formulation, Theorems 1 and 2 yield the value of the minimum achievable leakage in closed-form.

Next, we provide the minimum achievable leakage in closed-form with respect to Min-Entropy. We establish this result in two ways: first, as a direct corollary of Theorem 1 to showcase the versatility of our main result, and next, through a game theoretic argument which provides intuition about the optimal strategies, and in particular, provides the optimal guessing strategies of the adversary as well.

Proposition 4: For a given prior $\boldsymbol{P}$ and cloak size-cap k, the maximum achievable posterior entropy with respect to Min-Entropy is $-\log(\max(1/k, \boldsymbol{P}_{[1]}))$. This in turn implies that the minimum achievable leakage with respect to Min-Entropy, denoted by $L_\infty(\boldsymbol{P}, k)$, is as follows: $L_\infty(\boldsymbol{P}, k) = 0$ for any $k \ge 1/\boldsymbol{P}_{[1]}$, and $L_\infty(\boldsymbol{P}, k) = -\log(k\boldsymbol{P}_{[1]})$ if $k < 1/\boldsymbol{P}_{[1]}$.

Proof: From Theorem 1, if $\boldsymbol{P}_{[1]} \le 1/k$, then $J = 1$, which means the highest achievable posterior entropy is $H(\boldsymbol{\pi_1}) = H((1/k, \ldots, 1/k))$. For Min-Entropy, this gives $-\log(1/k)$. If on the other hand $\boldsymbol{P}_{[1]} > 1/k$, then J is an index between 2 and k. For any $J > 1$, the largest element of $\boldsymbol{\pi_J}$ is $\boldsymbol{P}_{[1]}$, hence $H(\boldsymbol{\pi_J})$ for Min-Entropy is equal to $-\log(\boldsymbol{P}_{[1]})$. Putting these together yields the claim. ∎

And now an alternative proof based on game theory:

Proof: In the 2PZSG corresponding to Min-Entropy, consider the following strategy of the adversary, which we refer to as $\boldsymbol{\alpha}_1$: always guess θ_1 for any observed M that includes θ_1, and make a uniformly random guess from M if it does not.[16] We argue that this strategy "guarantees" an expected reward of p_1 for the adversary, *irrespective* of the strategy of the defender. This simply follows by conditioning on the realization of the secret: with probability p_1, the secret is θ_1. For such a case, any feasible cloak that the defender chooses, the adversary will guess correctly and gains one unit (since any feasible cloak for θ_1 should include it as well). Now, consider this alternative strategy of the adversary, denoted by $\boldsymbol{\alpha}_2$: For any observed M, choose a uniformly random guess from it.[17] This strategy "guarantees" an expected reward of $1/k$ for the adversary, irrespective of the strategy of the defender. This can be seen by conditioning on the observed cloaks, the expected reward of the adversary that follows this strategy is $\sum_{M\in\mathcal{M}^+} \Pr(M) \sum_{\theta\in M} \boldsymbol{P}(\theta|M) \cdot \frac{1}{|M|} \ge \sum_{M\in\mathcal{M}^+} \Pr(M) \cdot \frac{1}{k} \cdot \sum_{\theta\in M} \boldsymbol{P}(\theta|M) = \sum_{M\in\mathcal{M}^+} \Pr(M) \cdot \frac{1}{k} = \frac{1}{k}$. Note in particular, that the specific induced posterior distributions do not play a role.

The adversary can guarantee an expected reward of p_1 and $1/k$ by employing the simple strategies $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ respectively. Hence, he can also guarantee the better of the two, i.e., $\max(1/k, p_1)$, just by comparing the values of p_1 and $1/k$ and employing the corresponding strategy. He may even be able to do better by perhaps employing more intelligent strategies. However, the defender can also guarantee that, irrespective of the strategy of the adversary, his gain never exceeds $\max(1/k, p_1)$ by adopting the strategies prescribed by Algorithm 1. In particular, if $p_1 \le 1/k$, then the defender has a feasible strategy that induces uniform posterior distribution for any chosen cloak.

[16]Formally, for all $M : \theta_1 \in M$, take $\boldsymbol{\alpha}_1(\theta_1; M) = 1$ and $\boldsymbol{\alpha}_1(\theta; M) = 0 \ \forall \theta \ne \theta_1$, and for all $M : \theta_1 \notin M$, take $\boldsymbol{\alpha}_1(\theta; M) = 1/|M| \ \forall \theta \in M$ and $\boldsymbol{\alpha}_1(\theta; M) = 0 \ \forall \theta \notin M$.

[17]Formally: $\boldsymbol{\alpha}(\theta; M) = 1/|M| \ \forall \theta \in M$ and $\boldsymbol{\alpha}(\theta; M) = 0 \ \forall \theta \notin M$, for all $M \in \mathcal{M}$.

Hence, irrespective of the guessing strategy of the adversary, his gain is going to be $1/k$. On the other hand, if $p_1 > 1/k$, then the defender has a feasible strategy that for any chosen cloak induces posterior distribution with its maximum value on θ_1 (so in part, for any realization of the secret, θ_1 is also picked as part of the cloak). This guarantees that irrespective of the strategy of the adversary, his gain is bounded by p_1. ■

Proposition (4) may come as a bit of surprise: if $\boldsymbol{P}_{[1]} > 1/k$, the information leakage with respect to Min-Entropy can be made absolutely zero. This can in fact be generalized to l-Guess-Entropy as well: if $\boldsymbol{P}_{[l]} \geq \frac{\sum_{i=l}^{k} \boldsymbol{P}_{[i]}}{}(k-l+1)$, then the minimum leakage with respect to l-Guess-Entropy is zero. These results however do not contradict the Shannon's perfect secrecy, since, unlike Shannon's entropy, Min-Entropy and l-Guess-Entropy do not retain the information of the whole distribution. Also note that these zero-leakage cases correspond to priors that are highly skewed. In such cases, figuratively speaking, the "giants" cannot be helped, but the secrets with small probabilities can "hide in the shadow of the giants". In other words, the prior already is very revealing and gives a lot of advantage to the adversary, but the defender can at least leverage those high probability secrets to not reveal any extra information to the adversary.

Extensions of Prop. 4 to the l-Guess and Guesswork are the following, which again can be seen as direct corollaries of Theorem 1 or established using game theoretic arguments.

Proposition 5: In the 2PZSG corresponding to the l-Guess-Error-Probability entropy, the maximum expected reward of the adversary is the following:

$$\max_{0\leq j\leq l}\left\{\sum_{i=1}^{j}\boldsymbol{P}_{[i]}+\big(1-\sum_{i=1}^{j}\boldsymbol{P}_{[i]}\big)\frac{l-j}{k-j}\right\}$$

Proposition 6: In the 2PZSG corresponding to the Guesswork entropy, the minimum expected cost of the adversary (expected number of guesses before detection) is:

$$\min_{1\leq j\leq k}\left\{\sum_{i=1}^{j-1}i\boldsymbol{P}_{[i]}+\big(1-\sum_{i=1}^{j-1}\boldsymbol{P}_{[i]}\big)\frac{k+j}{2}\right\}$$

VI. Numerical Illustrations

First, we investigate the effect of the maximum permissible cloak size, k, and the choice of the entropy measures on the minimum achievable leakage. We consider three candidate entropy measures: Shannon, Guesswork, and Min-Entropy. Recall that: $\mathcal{H}_{\mathrm{Sh.}}(\boldsymbol{\theta}) := -\sum_{i=1}^{n} p_i \log(p_i)$ and its conditional entropy follows the form of (4). For Min-Entropy, $\mathcal{H}_{\infty}(\boldsymbol{\theta}) := -\log\max_{\theta}(\boldsymbol{P}(\theta))$, and conditional entropy takes the form of (9) (a case of (2b)), that is, $\mathcal{H}_{\infty}(\boldsymbol{\theta}|\boldsymbol{M}) := -\log\sum_{M\in\mathcal{M}}\Pr(M)\max_{\theta}\boldsymbol{P}(\theta|M)$. For Guesswork, $\mathcal{H}_{\mathrm{Gu.}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} i\boldsymbol{P}_{[i]}$ and the posterior entropy follows the prescription of (4). However, to obtain a comparable scale for all three, we added a log to both prior and posterior of Guesswork entropy (hence, a case of (2a)).

For all examples in this section, we consider a secret space consisting of 30 elements with the following prior distribution: $\boldsymbol{P} = (30/465, 29/465, \ldots, 1/465)$. Fig. 1a shows that, as we expect, the minimum leakage reduces as larger cloaks are allowed. When leakage is quantified with Shannon entropy, min-leakage only vanishes when $k = n$, in accordance with the classic "perfect secrecy" result. However, the minimum achievable information leakage with respect to Min-Entropy becomes zero for any $k \geq 16$ in our example. This is in accordance with the result of Proposition 4, which stated that for any $k \geq \lceil 1/\boldsymbol{P}_{[1]}\rceil$, an optimal cloaking strategy can achieve zero leakage with respect to Min-Entropy. In this example, $\lceil 1/P_{[1]}\rceil = \lceil 465/30\rceil = 16$.

Next, we compare the performance of an optimal strategy against the following base-line strategy: *for any given secrets pick a feasible cloak uniformly randomly*. Note that this strategy is in fact optimal when the prior distribution is uniform, but not necessarily for other priors. Fig. 1b depicts the leakage with respect to Min-Entropy achieved by the optimal strategy and the base-line strategy when $\boldsymbol{P} = (30/465, \ldots, 1/465)$, demonstrating the sub-optimality of the base-line for any intermediate value of k. Adoption of this strategy is sub-optimal because it essentially ignores the fact that an adversary who has learned the distribution of the secret can exploit it to further improve his guessing power.

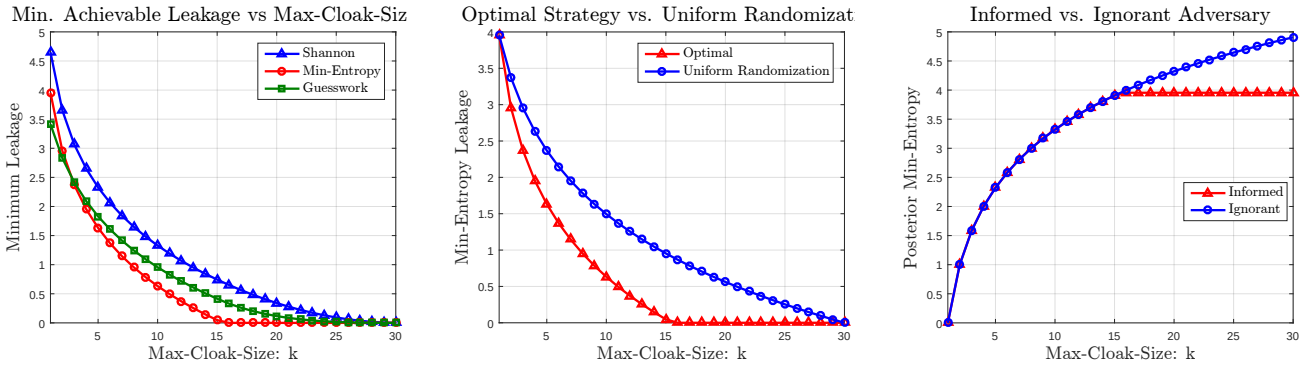
Next, we investigate the effect of one of the assumptions we made in the paper: that the defender designs her cloaking strategy assuming that the adversary knows the true distribution of the secrets. In particular, we consider an "uninformed" adversary, that does not know the prior distribution, and thus, for any observed cloak simply chooses a guess uniformly randomly. What will be the performance of the strategy that is designed to be optimal with the worst-case assumption that the adversary is "informed" of the true distribution, but facing an uninformed adversary. In Fig. 1c, for the prior of $\boldsymbol{P} = (30/465, \ldots, 1/465)$, we have depicted the posterior Min-Entropy for an "informed" vs. "ignorant" adversary. As we can see, for $k \geq 16$, the Min-Entropy of the ignorant adversary is larger than that of the informed one. For $k < 16$, we have $\boldsymbol{P}_{[1]} < 1/k$, which implies $J = 1$, and hence, the optimal cloaking strategy indeed induces uniform posterior distributions on any chosen cloak. Hence, uniformly random guessing from any observed cloak by the "uninformed" adversary is exactly optimal.

VII. Acknowledgment

The authors' work was supported by EPSRC grant EP/K005820/1.

VIII. Conclusion and Future Work

We investigated the fundamental properties of leakage when perfect secrecy is not achievable due to the limit on the allowable size of the pre-images of outputs. We have shown the existence of universally optimal strategies achieving minimum leakage for any measure of entropy that



(a) (b) (c)

Figure 1. For all figures, the prior is $\boldsymbol{P} = (30/465, \ldots, 1/465)$ and the max-cloak-size is varied from 1 to $n = 30$ as the x-axis. (a) The minimum achievable leakage with respect to Shannon, Guesswork and Min-Entropy, . Note that, as we expect, the minimum leakage improves as larger cloaks are allowed. The Shannon entropy only becomes zero for $k = n$, as is the classic perfect secrecy, while the best min-entropy leakage becomes zero for any $k \geq 16$ for this prior distribution as per Proposition 4. (b) Comparison of the Min-Entropy leakage achieved by the optimal strategy and the base-line (uniform randomization) strategy. (c) Negative of the log of the expected reward of an "informed" adversary, who knows the true prior distribution of the secret, and an "uninformed" adversary who simply assumes a uniform prior. The channel (randomized strategies) is designed to be optimal assuming facing an informed adversary.

satisfy a mild set of conditions (symmetry, expansibility, and what we called "core-concavity") and extended it to g-leakage entropies where these conditions may fail. We also demonstrated how the problem of minimum leakage channel design is equivalent to Nash equilibria in a corresponding two person zero-sum games of incomplete information for a range of entropy measures.

There are several possible directions to explore for future research. For instance, our theorems give rise to the next fundamental question: for what other types of constraints can we construct universally optimal channels? We expect that the techniques developed in our proofs be reused specially in unification of different notions of leakage and establishing robustness results. Another interesting direction of research is relaxing the i.i.d. assumption of the realization of the secrets and allowing correlation over time, where an adversary can learn more than just the distribution by observing the history. Designing optimal strategies for such a case is our next research goal.

References

[1] J. Heusser and P. Malacaria, "Quantifying information leaks in software," in *Twenty-Sixth Annual Computer Security Applications Conference, ACSAC 2010, Austin, Texas, USA, 6-10 December 2010*, 2010, pp. 261–269.

[2] G. Doychev, B. Köpf, L. Mauborgne, and J. Reineke, "Cacheaudit: a tool for the static analysis of cache side channels," *ACM Transactions on Information and System Security (TISSEC)*, vol. 18, no. 1, p. 4, 2015.

[3] A. McIver, C. Morgan, and T. M. Rabehaja, "Abstract hidden Markov models: A monadic account of quantitative information flow," in *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, Kyoto, Japan, July 6-10, 2015*, 2015, pp. 597–608.

[4] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *Computer Security Foundations Symposium (CSF), 2014 IEEE 27th*. IEEE, 2014, pp. 308–322.

[5] C. E. Shannon, "Communication theory of secrecy systems*," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[6] D. Clark, S. Hunt, and P. Malacaria, "Quantitative information flow, relations and polymorphic types," *Journal of Logic and Computation*, vol. 15, no. 2, pp. 181–199, 2005.

[7] G. Smith, "On the foundations of quantitative information flow," in *Foundations of software science and computational structures*. Springer, 2009, pp. 288–302.

[8] A. McIver, L. Meinicke, and C. Morgan, "Compositional closure for Bayes risk in probabilistic noninterference," in *Automata, Languages and Programming*. Springer, 2010, pp. 223–235.

[9] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring Information Leakage Using Generalized Gain Functions," in *Computer Security Foundations Symposium (CSF), 2012 IEEE 25th*. IEEE, 2012, pp. 265–279.

[10] A. McIver, C. Morgan, G. Smith, B. Espinoza, and L. Meinicke, "Abstract channels and their robust information-leakage ordering," in *Principles of Security and Trust*. Springer, 2014, pp. 83–102.

[11] B. Köpf and G. Smith, "Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks," in *Computer Security Foundations Symposium (CSF), 2010 23rd IEEE*. IEEE, 2010, pp. 44–56.

[12] B. Köpf and M. Durmuth, "A provably secure and efficient countermeasure against timing attacks," in *Computer Security Foundations Symposium, 2009. CSF'09. 22nd IEEE*. IEEE, 2009, pp. 324–335.

[13] M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for web transactions," *ACM Transactions on Information and System Security (TISSEC)*, vol. 1, no. 1, pp. 66–92, 1998.

[14] A. Khoshgozaran and C. Shahabi, "Private information retrieval techniques for enabling location privacy in location-based services," in *Privacy in Location-Based Applications*. Springer, 2009, pp. 59–83.

[15] C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. Di Vimercati, and P. Samarati, "Location privacy protection through obfuscation-based techniques," in *Data and Applications Security XXI*. Springer, 2007, pp. 47–60.

[16] A. Gervais, R. Shokri, A. Singla, S. Capkun, and V. Lenders, "Quantifying web-search privacy," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 966–977.

[17] P. Malacaria, "Algebraic foundations for quantitative information flow," *Mathematical Structures in Computer Science*, vol. 25, no. 2, pp. 404–428, 2015.

[18] J. Domingo-Ferrer, A. Solanas, and J. Castellà-Roca, "h (k)-private information retrieval from privacy-uncooperative queryable databases," *Online Information Review*, vol. 33, no. 4, pp. 720–744, 2009.

[19] G. Doychev and B. Köpf, "Rational protection against timing attacks," in *Computer Security Foundations Symposium (CSF), 2015 IEEE 28th*. IEEE, 2015, pp. 526–536.

[20] G. Theodorakopoulos, R. Shokri, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services," in *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, 2014, pp. 73–82.

[21] B. Köpf and D. Basin, "An information-theoretic model for adaptive side-channel attacks," in *Proceedings of the 14th ACM conference on Computer and communications security*. ACM, 2007, pp. 286–296.

[22] M. S. Alvim, M. E. Andrés, and C. Palamidessi, "Information flow in interactive systems," in *CONCUR 2010-Concurrency Theory*. Springer, 2010, pp. 102–116.

[23] M. Boreale and F. Pampaloni, "Quantitative information flow under generic leakage functions and adaptive adversaries," in *Formal Techniques for Distributed Objects, Components, and Systems*. Springer, 2014, pp. 166–181.

[24] F. Biondi, T. Given-Wilson, and A. Legay, "Attainable unconditional security for shared-key cryptosystems," in *The 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15)*, 2015.

[25] I. Csiszár, "Axiomatic characterizations of information measures," *Entropy*, vol. 10, no. 3, pp. 261–273, 2008.

[26] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[27] S. Fehr and S. Berens, "On the conditional rényi entropy," *Information Theory, IEEE Transactions on*, vol. 60, no. 11, pp. 6801–6810, 2014.

[28] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional rényi entropies," in *Information Theoretic Security*. Springer, 2014, pp. 103–121.

[29] M. Backes, G. Doychev, and B. Köpf, "Preventing Side-Channel Leaks in Web Traffic: A Formal Approach," in *Proc. 20th Network and Distributed Systems Security Symposium (NDSS)*, 2013.

[30] F. Biondi, A. Legay, P. Malacaria, and A. Wąsowski, "Quantifying information leakage of randomized protocols," in *Verification, Model Checking, and Abstract Interpretation*. Springer, 2013, pp. 68–87.

[31] P. Mardziel, M. S. Alvim, M. Hicks, and M. R. Clarkson, "Quantifying information flow for dynamic secrets," in *2014 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2014, pp. 540–555.

APPENDIX A.
PROOF OF THEOREM 1

We develop the proof in the following logical succession: first, we establish a parametric set of upper-bounds for the posterior entropy $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]$ for any cloaking strategy (Lemma 2). Then we show that our algorithm provides a feasible cloaking strategy that achieves one of such upper-bounds, and hence is optimal (Lemma 3).

Lemma 2: (part-I) For any $\boldsymbol{\delta} \in \mathscr{D}$, $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] \leq H(\boldsymbol{\pi_1})$. (part-II) Suppose that $p_{j-1} > \dfrac{\sum_{i=j}^{n} p_i}{k-j+1}$ for a $j \in \{2,\dots,k\}$. Then for any cloaking strategy $\boldsymbol{\delta} \in \mathscr{D}$, $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] \leq H(\boldsymbol{\pi_j})$.

Proof: Recall from (3) that the general form of the conditional entropy is as: $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] = \eta\big(\sum_{M\in\mathcal{M}^+} \Pr(M) F(\boldsymbol{P}(\boldsymbol{\theta}|M))\big)$ where g and F functions satisfy (2a) or (2b). We provide the proof for the case of (2a). The arguments for case (2b) follows similarly. For case (2a), since F is symmetric and concave, it is also Schur-concave.

Consider an arbitrary randomized cloaking strategy $\boldsymbol{\delta} \in \mathscr{D}$. Then for any $M \in \mathcal{M}^+$, we have $|\operatorname{supp}(\boldsymbol{P}(\boldsymbol{\theta}|M))| \leq k$, that is, at most k entries of $\boldsymbol{P}(\boldsymbol{\theta}|M)$ are non-zero. This is due to the facts that $|M| \leq k$ and $\boldsymbol{P}(\theta|M) = 0$ for any $\theta \notin M$. To see the latter, recall that any assigned cloak for a secret must include that secret, which lead to requirement (1b) for any $\boldsymbol{\delta} \in \mathscr{D}$, which in turn, along with (1a) implies $\boldsymbol{\delta}(M;\theta) = 0$ for all $\theta \notin M$.

For part-I, first recall that $\boldsymbol{\pi_1}$ is the uniform distribution over support size of k, which is majorized by any distribution over a support of at most size k. Therefore, following Schur-concavity of F, each term $F(\boldsymbol{P}(\boldsymbol{\theta}|M))$ is bounded

by $F(\boldsymbol{\pi_1})$. Hence, also using the fact that η is an increasing scalar function, we have:

$$\begin{aligned}\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] &= \eta\left(\sum_{M\in\mathcal{M}^+}\Pr(M)F(\boldsymbol{P}(\boldsymbol{\theta}|M))\right)\\ &\le \eta\left(\sum_{M\in\mathcal{M}^+}\Pr(M)F\left(\boldsymbol{\pi_1}\right)\right)\\ &= \eta\left(F\left(\boldsymbol{\pi_1}\right)\sum_{M\in\mathcal{M}^+}\Pr(M)\right) = \eta\left(F\left(\boldsymbol{\pi_1}\right)\right) = H(\boldsymbol{\pi_1})\end{aligned}$$

Part-I was quite intuitive: the highest uncertainty of the adversary, if the cloaks have to be restricted to size k, corresponds to uniform distribution over k items. We address part (II) next.

First, for each $M \in \mathcal{M}^+$, following the "symmetry" property of F, we have: $F(\boldsymbol{P}(\boldsymbol{\theta}|M)) = F(\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M))$, that is, we can safely sort each of the posterior probabilities in descending order.

Second, as we argued, $M \in \mathcal{M}^+$, $|\operatorname{supp}(\boldsymbol{P}(\boldsymbol{\theta}|M))| \le k$, which means that the bottom $n-k$ elements of $\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)$ are always zero. Therefore, following the "expansibility" property of F, we can safely remove them. That is, $F(\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)) = F(\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)})$, where the subscript $\downarrow(1,\ldots,k)$ denotes projecting to only the first k elements.

Third, we note the fact that the probability distribution $\Pr(M), M \in \mathcal{M}^+$ constitutes the coefficient of a convex combination: each is positive and they add up to one. Hence, following the "concavity" property of F and the previous two steps, we have:[18]

$$\begin{aligned}\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] &= \eta\left(\sum_{M\in\mathcal{M}^+}\Pr(M)F\left(\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)}\right)\right)\\ &\le \eta\left(F\left(\sum_{M\in\mathcal{M}^+}\Pr(M)\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)}\right)\right)\\ &= H\left(\sum_{M\in\mathcal{M}^+}\Pr(M)\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)}\right). \quad (12)\end{aligned}$$

Recall that $\boldsymbol{P}(\theta|M) = \boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\Pr(M)$ for an $M \in \mathcal{M}^+$. Therefore, letting $\boldsymbol{\delta}(M;\boldsymbol{\theta})$ denote the n-sized vector $(\boldsymbol{\delta}(M;\theta)), \theta \in \Theta$ and using the element-wise product symbol $\odot$, we can write: $\Pr(M)\boldsymbol{P}(\boldsymbol{\theta}|M) = \boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta})$. Therefore, the inequality in (12) can be re-written as: $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] \le H(\boldsymbol{Q})$ where $\boldsymbol{Q} = (q_i), i = 1,\ldots,k$ is defined as follows: $q_i := \sum_{M\in\mathcal{M}^+}(\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}$ (recall that subscript $[i]$ denotes the i'th largest element of a vector).

Fourth, we show that given the condition of the lemma, $\boldsymbol{Q}$ majorizes $\boldsymbol{\pi_j} = \left(p_1,\ldots,p_{j-1},\frac{\sum_{i=j}^n p_i}{k-j+1},\ldots,\frac{\sum_{i=j}^n p_i}{k-j+1}\right)$. First of all, $\boldsymbol{Q}$ is itself a probability distribution, since it is a convex combination of probability distributions $\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)}$. Hence, both $\boldsymbol{Q}$ and $\boldsymbol{\pi_j}$ are k-sized probability distributions and we have $\sum_{i=1}^k q_i = \sum_{i=1}^k \boldsymbol{\pi_j}(i) = 1$. Moreover, both $\boldsymbol{Q}$ and $\boldsymbol{\pi_j}$ are already in descending order: For $\boldsymbol{Q}$ this follows from the fact that all $\boldsymbol{P}^{\downarrow}(\boldsymbol{\theta}|M)_{\downarrow(1,\ldots,k)}$ were in descending order; for $\boldsymbol{\pi_j}$ as described in (7), this follows from the condition of the lemma that $p_{j-1} > \sum_{i=j}^n p_i/(k-j+1)$, along with the fact that $\boldsymbol{P}$ was in descending order, and hence, so is $(p_1,\ldots,p_{j-1})$. Therefore, all we need to show is that $\sum_{i=1}^l q_i \ge \sum_{i=1}^l \boldsymbol{\pi_j}(i)$ for all $l = 1,\ldots,(k-1)$. We will use the following sub-lemma:

Sub-lemma 1: $\sum_{i=1}^l q_i \ge \sum_{i=1}^l p_i$ for any $l < k$.

Proof: Recall $q_i := \sum_{M\in\mathcal{M}^+}(\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}$ from step 3. Hence, for any $l < k$:

$$\begin{aligned}\sum_{i=1}^l q_i &= \sum_{i=1}^l\sum_{M\in\mathcal{M}^+}(\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}\\ &= \sum_{M\in\mathcal{M}^+}\sum_{i=1}^l(\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]} \ge \sum_{M\in\mathcal{M}^+}\sum_{i=1}^l\boldsymbol{P}(\theta_i)\boldsymbol{\delta}(M;\theta_i)\end{aligned}$$

The last inequality follows because summation of the top l elements of any vector is no less than the summation of any l elements of it. The right hand side of the inequality, after a change in the order of summations, is equal to: $\sum_{i=1}^l\sum_{M\in\mathcal{M}^+}\boldsymbol{P}(\theta_i)\boldsymbol{\delta}(M;\theta_i) = \sum_{i=1}^l p_i\sum_{M\in\mathcal{M}^+}\boldsymbol{\delta}(M;\theta_i) = \sum_{i=1}^l p_i$. The last equality follows because any $\boldsymbol{\delta}\in\mathscr{D}$ satisfies $\sum_{M\in\mathcal{M}^+}\boldsymbol{\delta}(M;\theta) = 1$ for each $\theta\in\Theta$. Replacing this back in the inequality yields the statement of the sub-lemma. ■

Now, for any $l \le j-1$, the inequality $\sum_{i=1}^l q_i \ge \sum_{i=1}^l \boldsymbol{\pi_j}(i)$ follows directly from the sub-lemma, since $\boldsymbol{\pi_j}(i) = p_i$ for all $i \le j-1$, by its definition in (7). For an $l \in \{j,\ldots,k-1\}$, first we argue that $\sum_{i=j}^l q_i/(l-j+1) \ge \sum_{i=j}^k q_i/(k-j+1)$: The left hand side is the (arithmetic) average of $(q_j,\ldots,q_l)$, and the right hand side is the (arithmetic) average of $(q_j,\ldots,q_k)$; the inequality then follows due to the fact that q_i's are in descending order. This inequality can be written as $\sum_{i=j}^l q_i \ge \frac{l-j+1}{k-j+1}\sum_{i=j}^k q_i$. Adding $\sum_{i=1}^{j-1} q_i$ to both sides, and rewriting $\sum_{i=j}^k q_i$ equivalently as $(1-\sum_{i=1}^{j-1} q_i)$, we obtain: $\sum_{i=1}^l q_i \ge \sum_{i=1}^{j-1} q_i + \frac{l-j+1}{k-j+1}(1-\sum_{i=1}^{j-1} q_i)$. Following Sub-lemma 1, we have $\sum_{i=1}^{j-1} q_i \ge \sum_{i=1}^{j-1} p_i$. Now, consider the scalar function $f(x) = x + \frac{l-j+1}{k-j+1}(1-x)$. For any $j\in\{2,\ldots,k\}$, this function is increasing in x. Therefore, $\sum_{i=1}^{j-1} q_i \ge \sum_{i=1}^{j-1} p_i$ implies $\sum_{i=1}^{j-1} q_i + \frac{l-j+1}{k-j+1}(1-\sum_{i=1}^{j-1} q_i) \ge \sum_{i=1}^{j-1} p_i + \frac{l-j+1}{k-j+1}(1-\sum_{i=1}^{j-1} p_i)$ as well. Note that the right hand side of the latter inequality is exactly $\sum_{i=1}^l \boldsymbol{\pi_j}(i)$ when $j-1 \le l \le k$. Putting the inequalities together, we obtain $\sum_{i=1}^l q_i \ge \sum_{i=1}^l \boldsymbol{\pi_j}(i)$ for any $l \in j,\ldots,k$. This completes the argument for establishing the claim that whenever the condition of the lemma is satisfied, we have $\boldsymbol{Q} \succ \boldsymbol{\pi_j}$, i.e., $\boldsymbol{Q}$ majorizes $\boldsymbol{\pi_j}$.

[18] Or equivalently, applying Jensen's inequality.

In the final step, we note that "Schur-concavity" of H together with $\boldsymbol{Q} \succ \boldsymbol{\pi_j}$ give $H(\boldsymbol{Q}) \leq H(\boldsymbol{\pi_j})$. The lemma now follows by noting that in step 3, we showed $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}] \leq H(\boldsymbol{Q})$. ∎

Lemma 3: For a given $\boldsymbol{P}$ and k, Algorithm 1: (I) produces a strategy $\boldsymbol{\delta}$, (II) that is feasible, and (III) that is optimal.

Proof: Part (I): We need to show that Algorithm 1 indeed terminates with an output $\boldsymbol{\delta}$. As we argued after (6), J can always be found. Therefore, the only step of the algorithm that we must ensure returns a result is finding an all positive solution to the linear system in step 2, which we prove next.

Sub-lemma 2: Given $\boldsymbol{P}$ (in descending order) and k, if $p_J \leq \frac{\sum_{i=J}^{n} p_i}{k-J+1}$ where $J \leq k$, then the following system has a solution: $\sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M = p_i,\ \forall i = J,\ldots,n$; subject to: $x_M \geq 0\ \forall M \in \mathcal{M}^*(\{1,\ldots,J-1\})$. Moreover, any solution satisfies: $\sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1\})} x_M = \frac{\sum_{i=J}^{n} p_i}{k-J+1}$.

Proof: For brevity, take $s := \sum_{i=J}^{n} p_i$, $t := (n-J+1)$, and $w := (k-J+1)$. Let $\Omega := \{\boldsymbol{y} = (y_i) \in \mathbb{R}^t : \sum_{i=1}^{t} y_i = s,\ \textit{and}\ 0 \leq y_i \leq s/w, \forall i = 1,\ldots,t\}$. Ω is a *convex polyhedron* in $\mathbb{R}^t$ (since it is described by a system of linear inequalities). It is also closed (clearly), and is non-empty, as $(s/t,\ldots,s/t) \in \Omega$. Hence, Ω is also a non-empty *polytope* in $\mathbb{R}^t$ (i.e., can be described as the convex hull of a finite number of points in $\mathbb{R}^t$). Specifically, any point inside Ω can be written as a convex combination of the *extreme* (a.k.a. *corner*) points of Ω (and vice versa).[19] The extreme points of Ω are t-dimensional vectors where w of their elements are s/w and the $t-w$ rest of them are zeros. There are $\binom{t}{w}$ of such vectors. Let Λ be a matrix whose columns are these extreme points, i.e, each column is a distinct permutations of w entries of s/w and $t-w$ entries of zero, that is: $\Lambda := [(s/w,\ldots,s/w,0,\ldots,0)^T;\ldots;(0,\ldots,0,s/w,\ldots,s/w)^T]$.

Now, let $\boldsymbol{q} := (p_J,\ldots,p_n)$. By construction, the condition of the sub-lemma on $\boldsymbol{P}$ can be equivalently stated as $\boldsymbol{q} \in \Omega$. Hence, $\boldsymbol{q}$ can be expressed as a convex combination of the extreme points of Ω. Let $\boldsymbol{z} \in \mathbb{R}^{+\binom{t}{w}}$ denote such a convex combination, thus, we have: $\Lambda \boldsymbol{z} = \boldsymbol{q}$ where $\boldsymbol{z} \geq \boldsymbol{0}$ (element-wise positive), and $\boldsymbol{1}^T \boldsymbol{z} = 1$.

On the other hand, the linear system in the sub-lemma can be written in matrix form as: $\Lambda' \boldsymbol{x} = \boldsymbol{q}$ where Λ' is a $t \times \binom{t}{w}$ matrix whose columns are all the $\binom{t}{w}$ permutations of having w entries of 1 and $t-w$ entries of 0. Therefore, with some re-ordering of the equations if necessary, we can write: $\Lambda = (s/w)\Lambda'$. Hence, $\Lambda \boldsymbol{z} = \boldsymbol{q}$ implies $(s/w)\Lambda' \boldsymbol{z} = \boldsymbol{q}$, and $\boldsymbol{z} \geq \boldsymbol{0}$ implies $(s/w)\boldsymbol{z} \geq \boldsymbol{0}$. Therefore, $\boldsymbol{x} = (s/w)\boldsymbol{z}$ is a feasible solution of the system in the sub-lemma.

The second claim of the sub-lemma follows from summating all the equations of the system and a simple counting: $\sum_{i=J}^{n} p_i = \sum_{i=J}^{n} \sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M = (k-J+1) \sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1\})} x_M$. ∎

[19] In fact, according to *Carathèodory's theorem*, this can be done by a convex combination of at most $t+1$ of them.

Note that the condition of the sub-lemma is clearly satisfied for J found in the first line of Algorithm 1.

Proof: For brevity, take $s := \sum_{i=J}^{n} p_i$, $t := (n-J+1)$, and $w := (k-J+1)$. Let $\Omega := \{\boldsymbol{y} = (y_i) \in \mathbb{R}^t : \sum_{i=1}^{t} y_i = s,\ \textit{and}\ 0 \leq y_i \leq s/w, \forall i = 1,\ldots,t\}$. Ω is a *convex polyhedron* in $\mathbb{R}^{t-1}$ (since it is described by a system of linear inequalities. The minus 1 is due to the one equality constraint). It is also closed (clearly), and is non-empty, as $(s/t,\ldots,s/t) \in \Omega$. Hence, Ω is also a non-empty *polytope* in $\mathbb{R}^{t-1}$ (i.e., can be described as the convex hull of a finite number of points in $\mathbb{R}^{t-1}$). Specifically, any point inside Ω can be written as a convex combination of the *extreme* (a.k.a. *corner*) points of Ω (and vice versa). In fact, according to *Carathèodory's theorem*, this can be done by a convex combination of at most t of them. The extreme points of Ω are t-dimensional vectors where w of their elements are s/w and the $t-w$ rest of them are zeros. There are $\binom{t}{w}$ of such vectors. Let Λ be a matrix whose columns are these extreme points, i.e, each column is a distinct permutations of w entries of s/w and $t-w$ entries of zero, that is: $\Lambda := [(s/w,\ldots,s/w,0,\ldots,0)^T;\ldots;(0,\ldots,0,s/w,\ldots,s/w)^T]$.

Now, let $\boldsymbol{q} := (p_J,\ldots,p_n)$. By construction, the condition of the sub-lemma on $\boldsymbol{P}$ can be equivalently stated as $\boldsymbol{q} \in \Omega$. Hence, $\boldsymbol{q}$ can be expressed as a convex combination of the extreme points of Ω. Let $\boldsymbol{z} \in \mathbb{R}^{+\binom{t}{w}}$ denote such a convex combination, thus, we have: $\Lambda \boldsymbol{z} = \boldsymbol{q}$ where $\boldsymbol{z} \geq \boldsymbol{0}$ (element-wise positive), and $\boldsymbol{1}^T \boldsymbol{z} = 1$.

On the other hand, the linear system in the sub-lemma can be written in matrix form as: $\Lambda' \boldsymbol{x} = \boldsymbol{q}$ where Λ' is a $t \times \binom{t}{w}$ matrix whose columns are all the $\binom{t}{w}$ permutations of having w entries of 1 and $t-w$ entries of 0. Therefore, with some re-ordering of the equations if necessary, we can write: $\Lambda = (s/w)\Lambda'$. Hence, $\Lambda \boldsymbol{z} = \boldsymbol{q}$ implies $(s/w)\Lambda' \boldsymbol{z} = \boldsymbol{q}$, and $\boldsymbol{z} \geq \boldsymbol{0}$ implies $(s/w)\boldsymbol{z} \geq \boldsymbol{0}$. Therefore, $\boldsymbol{x} = (s/w)\boldsymbol{z}$ is a feasible solution of the system in the sub-lemma.

The second claim of the sub-lemma follows from summating all the equations of the system and a simple counting: $\sum_{i=J}^{n} p_i = \sum_{i=J}^{n} \sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M = (k-J+1) \sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1\})} x_M$. ∎

Part (II): Next, we show that the returned solution $\boldsymbol{\delta}$ corresponds to a feasible randomized cloaking strategy. The conditions $\boldsymbol{\delta}(M;\theta) \geq 0$ are trivially satisfied by construction. Hence, we only need to show that for any $\theta \in \Theta$, $\sum_{M \in \mathcal{M}(\theta)} \boldsymbol{\delta}(M;\theta) = 1$. For an $i \in \{J,\ldots,n\}$, Algorithm 1 assigns $\boldsymbol{\delta}(M;i) = x_M/p_i$ for all $M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})$, and zero for any other M. Hence, $\sum_{M \in \mathcal{M}(i)} \boldsymbol{\delta}(M;i) = \sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M/p_i = 1$, where the last equality follows directly from the system of equations in Step 2 of the algorithm, specifically, the equality constraint of $\sum_{M \in \mathcal{M}^*(\{1,\ldots,J-1,i\})} x_M = p_i$. On the other hand, for an $i \in \{1,\ldots,J-1\}$, the algorithm assigns: $\boldsymbol{\delta}(M;i) = x_M(k-J+1)/\sum_{j=J}^{n} p_j$ for all $M \in \mathcal{M}^*(\{1,\ldots,J-1\})$ and zero for any other M. There-

fore, $\sum_{M\in\mathcal{M}(i)}\boldsymbol{\delta}(M;i)=\sum_{M\in\mathcal{M}^*(\{1,\ldots,J-1\})}x_M(k-J+1)/\sum_{j=J}^{n}p_j=1$, where the last equality is due to the second claim of Sub-lemma 2, that $\sum_{M\in\mathcal{M}^*(\{1,\ldots,J-1\})}x_M=(\sum_{i=J}^{n}p_i)/(k-J+1)$.

Part (III): Finally, we show that the reported feasible strategy is indeed optimal. First, note that only cloaks M for which $x_M>0$ receive a non-zero probability. All of such cloaks are in $\mathcal{M}^*(\{1,\ldots,J-1\})$, i.e., include secrets $\theta_1,\ldots,\theta_{J-1}$ along with $k-J+1$ others. Let $M=\{\theta_1,\ldots,\theta_{J-1},\phi_1,\ldots,\phi_{k-J+1}\}$, where $\{\phi_1,\ldots,\phi_{k-J+1}\}\subset\{\theta_J,\ldots,\theta_n\}$ be any of such cloaks for which $x_M>0$. The posterior probability distribution for M (the updated probability distribution given M is observed) can be computed using the Bayes' rule: $\boldsymbol{P}(\theta|M)=\boldsymbol{P}(\theta)\boldsymbol{\delta}(M;\theta)/\Pr(M)$ where $\Pr(M)=\big(\sum_{\theta'\in M}\boldsymbol{P}(\theta')\boldsymbol{\delta}(M;\theta')\big)$. Replacing from the assignments in lines 3 through 5 of Algorithm 1, first, for $\Pr(M)$ we get:

$$\Pr(M)=\sum_{i=1}^{J-1}p_i\left(\frac{x_M(k-J+1)}{\sum_{j=J}^{n}p_j}\right)+\sum_{i=1}^{k-J+1}\boldsymbol{P}(\phi_i)\frac{x_M}{\boldsymbol{P}(\phi_i)}$$
$$=x_M(k-J+1)\left(\frac{\sum_{i=1}^{J-1}p_i}{\sum_{j=J}^{n}p_j}+1\right)=\frac{x_M(k-J+1)}{\sum_{j=J}^{n}p_j}$$

Hence, for all $i=1,\ldots,k-J+1$:

$$\boldsymbol{P}(\phi_i|M)=\frac{\boldsymbol{P}(\phi_i)x_M/\boldsymbol{P}(\phi_i)}{x_M(k-J+1)/\sum_{j=J}^{n}p_j}=\frac{\sum_{j=J}^{n}p_j}{k-J+1}$$

On the other hand, for $i=1,\ldots,J-1$:

$$\boldsymbol{P}(\theta_i|M)=\frac{\boldsymbol{P}(\theta_i)x_M(k-J+1)/\sum_{j=J}^{n}p_j}{x_M(k-J+1)/\sum_{j=J}^{n}p_j}=p_i$$

Hence, our randomized cloaking strategy ensures that for each $M\in\mathcal{M}^+$, $\boldsymbol{P}(\boldsymbol{\theta}|M)=\boldsymbol{\pi_J}$. This gives $\mathcal{H}[\boldsymbol{\theta}|\boldsymbol{M}]=H(\boldsymbol{\pi_J})$. Now, if $J=1$, then our algorithm achieves the upper-bound of Lemma 2-Part (I). If $J\geq 2$, then following its definition, we must have: $p_{J-1}>\sum_{i=J-1}^{n}p_i/(k-J+2)$. Multiplying both sides by $k-J+2$ and canceling out one of the p_J's, we get the equivalent inequality that $p_{J-1}>\sum_{i=J}^{n}p_i/(k-J+1)$. This is exactly the condition for Part (II) of Lemma 2. Hence, our randomized cloaking strategy achieves the upper-bound of $H(\boldsymbol{\pi_J})$. This concludes the proof of Lemma 3, and thus, of the theorem. ∎

Appendix B.
Proof of Theorem 2

The proof follows almost identically the steps of the proof of Theorem 1 with $\boldsymbol{P}$ replaced with $G\boldsymbol{P}$. As before, we first establish that $H(\boldsymbol{\pi_{g,J_g}})$ is among a set of upper-bounds for $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}]$ for any cloaking strategy, and then we prove that if Algorithm 1 is fed with $G\boldsymbol{P}=(\boldsymbol{\gamma}\odot\boldsymbol{P})$, sorted in descending order, produces a feasible strategy that achieves this upper-bound. The main observation to bear in mind is that although the entropy function $H_g(\boldsymbol{P})=H(G\boldsymbol{P})$ may no longer be symmetric, expansible or core-concave with respect to $\boldsymbol{P}$, but it has all these properties with respect to $G\boldsymbol{P}$ as assumed for the $H(\cdot)=\eta(F(\cdot))$ function. For brevity, we will only point out the important changes from the proof of Theorem 1.

Lemma 4: (part-I) For any $\boldsymbol{\delta}\in\mathscr{D}$, $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}]\leq H(\boldsymbol{\pi_{g,1}})$. (part-II) Suppose that $\gamma_{j-1}p_{j-1}>\frac{\sum_{i=j}^{n}\gamma_ip_i}{k-j+1}$ for a $j\in\{2,\ldots,k\}$. Then for any $\boldsymbol{\delta}\in\mathscr{D}$, $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}]\leq H(\boldsymbol{\pi_{g,j}})$.

Proof: The proof mirrors that of Lemma 2 where $\boldsymbol{P}$ is replaced by $G\boldsymbol{P}$, heeding the following two observations:

(1) For both parts, we used the fact that for any observed cloak, $|\operatorname{supp}(\boldsymbol{P}(\boldsymbol{\theta}|M))|\leq k$, since the secret is certainly one of the members of the cloak. For a diagonal matrix G, we have $\operatorname{supp}(G\boldsymbol{P}(\boldsymbol{\theta}|M))\subseteq\operatorname{supp}(\boldsymbol{P}(\boldsymbol{\theta}|M))$, since any zero entry of $\boldsymbol{P}$ leads to a corresponding zero entry in $G\boldsymbol{P}$. Therefore, $|\operatorname{supp}(\boldsymbol{P}(\boldsymbol{\theta}|M))|\leq k$ implies $|\operatorname{supp}(G\boldsymbol{P}\boldsymbol{\theta}|M))|\leq k$ too.

(2) Following the steps as in Lemma 2 leads to the inequality $\mathcal{H}_g[\boldsymbol{\theta}|\boldsymbol{M}]\leq H(\boldsymbol{Q})$ where $\boldsymbol{Q}=(q_i),i=1,\ldots,k$ is now defined as follows: $q_i:=\sum_{M\in\mathcal{M}^+}((G\boldsymbol{P})\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}$, which for $G=\operatorname{diag}(\boldsymbol{\gamma})$, is equal to $\sum_{M\in\mathcal{M}^+}(\boldsymbol{\gamma}\odot\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}$. Note that neither $\boldsymbol{Q}$ nor $\boldsymbol{\pi_{g,j}}$ are necessarily probability distributions. In order to show that $\boldsymbol{Q}$ still majorizes $\boldsymbol{\pi_{g,j}}$, following the definition of majorization, we first need to show $\sum_{i=1}^{k}q_i=\sum_{i=1}^{k}\boldsymbol{\pi_{g,j}}(i)$. For the first summation, we have:

$$\sum_{i=1}^{k}q_i=\sum_{i=1}^{k}\sum_{M\in\mathcal{M}^+}(\boldsymbol{\gamma}\odot\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}=$$
$$\sum_{M\in\mathcal{M}^+}\sum_{i=1}^{k}(\boldsymbol{\gamma}\odot\boldsymbol{P}\odot\boldsymbol{\delta}(M;\boldsymbol{\theta}))_{[i]}=\sum_{M\in\mathcal{M}^+}\sum_{i=1}^{k}\gamma_ip_i\boldsymbol{\delta}(M;\theta_i)$$
$$=\sum_{i=1}^{k}\gamma_ip_i\sum_{M\in\mathcal{M}^+}\boldsymbol{\delta}(M;\theta_i)=\sum_{i=1}^{k}\gamma_ip_i$$

On the other hand, replacing from the definition of $\boldsymbol{\pi_{g,j}}$ in the statement of the theorem, we get: $\sum_{i=1}^{k}\boldsymbol{\pi_{g,j}}(i)=\sum_{l=1}^{j-1}\gamma_lp_l+(k-j+1)\sum_{i=j}^{n}\gamma_ip_i/(k-j+1)=\sum_{i=1}^{k}\gamma_ip_i$, hence the first step for showing majorization is established.

For the second step, note that, as before, both $\boldsymbol{Q}$ and $\boldsymbol{\pi_{g,j}}$ are already in descending order: $\boldsymbol{Q}$ by design, and $\boldsymbol{\pi_{g,j}}$ due to the assumption (without loss of generality) that elements of $G\boldsymbol{P}$ are in descending order along with the inequality condition of the lemma. Hence we are only left to show that the partial sums of $\boldsymbol{Q}$ are larger than those of $\boldsymbol{\pi_{g,j}}$. This follows from a generalization of Sub-lemma 1. The only part that is different in its proof is to consider the function $f(x)=x+\frac{l-j+1}{k-j+1}(c-x)$, where $c=\sum_{i=1}^{n}\gamma_ip_i$, and noting that for any c, the function f is still increasing in x. ∎

The rest of the proof follows identically as in the proof of Theorem 1 if we replace $\boldsymbol{P}$ with $G\boldsymbol{P}$.

APPENDIX C.
PROOF OF PROPOSITION 3

Recall that $\mathcal{M}^+(\boldsymbol{\delta})$ denotes the set of cloaks that has a strictly positive probability of being observed by the adversary. For pure strategy profiles (and following the assumption that $\text{supp}(\boldsymbol{P}) = \Theta$), we have $\mathcal{M}^+(\boldsymbol{\delta}) = \boldsymbol{d}[\Theta]$, where $\boldsymbol{d}[\Theta]$ represents the image of the pure strategy of the defender, i.e., $\boldsymbol{d}[\Theta] := \{M \in \mathcal{M} | \exists \theta \in \Theta, M = \boldsymbol{d}(\theta)\}$.

Suppose the claim is false and there is a (pure) strategy profile $(\boldsymbol{d}^*, \boldsymbol{a}^*) \in \mathcal{D} \times \mathcal{A}$ that constitutes a (Nash) equilibrium. First, since the trivial cloak of the whole secret space is not permissible for all secrets, following the *pigeon-hole principle*, there must exist $M_1, M_2 \in \boldsymbol{d}^*[\Theta]$, $M_1 \neq M_2$ (recall that any cloak picked for a secret must contain it).

Let $\boldsymbol{a}^*(M_1) = \theta_1$ and $\boldsymbol{a}^*(M_2) = \theta_2$. We must have: $\boldsymbol{d}^*(\theta_1) = M_1$ and $\boldsymbol{d}^*(\theta_2) = M_2$, i.e., the strategy of the defender must choose M_1 for θ_1 and M_2 for θ_2. This follows from the fact that $\boldsymbol{a}^*$ must be a best response to $\boldsymbol{d}^*$. If for instance $\boldsymbol{d}^*(\theta_1) \neq M_1$, then, everything else fixed, the adversary can strictly improve his expected payoff by changing $\boldsymbol{a}^*(M_1)$ from θ_1 to any secret that chooses cloak M_1, i.e., any $\theta \in \boldsymbol{d}^{*-1}(M_1)$, where $\boldsymbol{d}^{*-1}(M_1)$ is the *pre-image*[20] of M_1 under $\boldsymbol{d}^*$ (note that $\boldsymbol{d}^{*-1}(M_1) \neq \emptyset$ because $M_1 \in \boldsymbol{d}^*[\Theta]$). This is because for a diagonal gain matrix with positive rewards, any non-zero probability of a correct guess strictly dominates a certain wrong guess for the adversary. Moreover, we must have: $\theta_1 \neq \theta_2$. To see this, suppose $\theta_1 = \theta_2$. Then, as we argued, we must both have $\boldsymbol{d}^*(\theta_1) = M_1$ and $\boldsymbol{d}^*(\theta_1) = M_2$, which cannot be, given that $\boldsymbol{d}^*$ is a legitimate pure strategy of the defender.

Consider the sets of permissible cloaks for θ_1 and θ_2, i.e., $\mathcal{M}(\theta_1)$ and $\mathcal{M}(\theta_2)$, respectively. Since $\boldsymbol{d}^*$ also is a best response to $\boldsymbol{a}^*$, we must have $\forall M \in \mathcal{M}(\theta_1)$: $\boldsymbol{a}^*(M) = \theta_1$, and $\forall M \in \mathcal{M}(\theta_2)$: $\boldsymbol{a}^*(M) = \theta_2$. This is because if, e.g., there exists an $M' \in \mathcal{M}(\theta_1)$ that $\boldsymbol{a}^*(M') \neq \theta_1$, then, everything else fixed, the defender has a strictly advantageous deviation from $\boldsymbol{d}^*(\theta_1) = M_1$ to $\boldsymbol{d}^*(\theta_1) = M'$. This is because for any sensible gain matrix, given the secret, a wrong guess of the adversary yields a strictly higher payoff for the defender than a correct guess. Now, following the "permissible connectivity" property, we have $\mathcal{M}(\theta_1) \cap \mathcal{M}(\theta_2) \neq \emptyset$. However, for any $M \in \mathcal{M}(\theta_1) \cap \mathcal{M}(\theta_2)$, we reach the contradicting requirement that $\boldsymbol{a}^*(M) = \theta_1$ and $\boldsymbol{a}^*(M) = \theta_2$ where $\theta_1 \neq \theta_2$. Therefore, $(\boldsymbol{d}^*, \boldsymbol{a}^*)$ could not have been an equilibrium.

[20] $\boldsymbol{d}^{-1}(M) := \{\theta \in \Theta | \boldsymbol{d}(\theta) = M\}$.