

Leakage-Minimal Design: Universality, Limitations, and Applications

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—We consider a setting where a system has to interact, and hence create distinct outputs (observables), but subject to such operational constraints wants to minimize the leakage that such observables reveal about its secret input. It has been previously demonstrated that under some (highly symmetrical) constraints on the observables, it is possible to design systems that are universally optimal in the sense of leaking minimal information no matter how information is measured.

In this work we make several contribution to this field. On universal (i.e., measure-invariant) optimality, we show its limitations through a counterexample where symmetry constraints are broken. Nevertheless, we also show two new universal optimality results: the first is in the presence of “graph like” constraints (that may lack symmetry). The second is universal optimality in the case of uncertainty about the prior. Furthermore, we prove that a generic class of leakage optimisation problems are convex problem, from which we derive that KKT conditions are necessary and sufficient for optimality. We demonstrate the practical value of the theory in the form of an application to timing attacks countermeasures.

I. INTRODUCTION

A common setting where privacy, anonymity, confidentiality, or secrecy is concerned, can be abstracted as a system that has an internal secret and produces publicly observable outputs. The realization of the secret is only known by the system and the secrecy goal is to leak the least information about the secret to potentially adversarial observers. One trivial way to achieve “zero-leakage” is to always exhibit the same observable irrespective of the realization of the secret. What makes the problem non-trivial is the fact that this trivial solution is indeed operationally unacceptable.

In secret communications, for instance, the goal is to transfer information to intended recipients. But producing the same observable, although guaranteeing zero leakage, will clearly mean no information transfer either. In the classical work of Shannon on secrecy [1], he proved that the only way to ensure zero leakage and have secret communication is to use “one-time-pads (OTP)” with the length at least as long as the secret, which guarantees the observables are completely independent from the secret to any observer other than the intended recipients, which hold the same copy of the OTP.

However, in many contexts other than secret communication where secrecy is a desirable goal, the trivial solution of producing the same observable for all secrets may be unsuitable or even infeasible. Consider, for instance, the system of a password-checker, where the password as the

secret: at the bare minimum, the system should produce two distinct observables (match/mismatch) to preserve its defining functionality. There are some other cases where zero leakage may be possible but undesirable as it leads to an unacceptable degradation in the utility of the system. Some prominent examples are:

- *Defence against timing side-channel channel*: Adversaries can gain information about the system by observing the time it takes a system to execute each process. Specially in the context of cryptography, these fluctuations in computation times can leak information about the secret key (an example of side channel attack) [2], [3]. One approach to remove this leakage is to release each of the computations only after a constant duration. This will guarantee zero leakage through the computation time side channel, as there will be only one observation irrespective of the secret. However, this may introduce huge delays as the computation time of each process is tied to the worst case scenario.
- *Defence against web-traffic fingerprinting*: Similar to bucketing, researchers have shown that adversaries can classify the type of encrypted web-traffic only by investigating the attributes of their traffic, most notably, the inter-packet-delays, as different types of web-traffic produce different “burstiness” patterns [4], [5]. Again, a solution can be devised that generates a single type of traffic, by a combination of delaying packets and releasing them only at certain intervals, and by generating dummy packets between moments of silence. Clearly, this zero-leakage remedy can lead to unacceptable overheads in delay and/or bandwidth.
- *Defence against device fingerprinting*: Similarly, an adversary can gain information about the device/OS/Browser through observing the structure of the request messages that they send to a web server [6]. Here, too, the zero-leakage solution may be undesirable/impossible to implement, as each browser uses different message headers or plug-ins for important purposes, especially “debuggability”.
- *Location Privacy*: Mobile users can take advantage of many “location-based-services”. The zero-leakage solution means that none of these utilities can be used, which could be an unacceptable level of degradation [7]–[9].

In each of these settings, lowering “information leaking” about the secret while respecting the operational constraints and overheads of the system is of interest. We develop a

unifying framework by sufficiently abstracting the underlying setting that capture the common theme of constrained secrecy. In particular, in all of these settings we can consider a system that, affected by an internal secret, generates observable outputs. Considering the secret and the observable as random variables, the system can be modelled as a “channel”, which is, a conditional distribution across observables given the secret. That is, given the realization of the secret, the system (potentially probabilistically) produces a particular observable. This is an equivalent interpretation of the system as a “strategy”, that is, a recipe of (potentially randomized) action per each secret. Throughout the paper, we will use “channels” and “strategies” interchangeably.

The problem then becomes the following: design a channel (strategy) such that: (1) it has minimal leakage, (2) none of the secrets are assigned to an unacceptable observable (“hard constraints”); (3) the “average” quality of performance of the system does not degrade too much (soft constraints);

But how should we quantify leakage to be able to compare among designs/channels/strategies? Leakage of information for channels has been studied by researchers in quantitative information flow [3], [10]–[15]. A candidate for measuring the amount of leakage was the difference between the posterior and prior Shannon entropies, however, it was shown that Shannon entropy may not be a good representative of the situation for some contexts. For instance, in password guessing attacks, more than the Shannon entropy, which relates to an adversary that can ask set-membership questions, the relevant measure of entropy is “Guesswork”, which is related to the expected number of guesses before striking on the correct one.

This raises a question of which entropy function to choose, given that they have different operational interpretation some of which depends on modelling the adversary. A recent work [16] showed that in a special case, one can design channels that are optimal no matter what measure of leakage is chosen and hence are in that sense “robust” against what adversary model the measure reflects.

In this work we explore this concept of universality, show the limitations of the previous results, establish universal properties and discuss a potential application.

Road-map and Contributions: The paper starts (Section II) by introducing the technical background relevant to this work. We will then prove (Section III) a fundamental result showing that the general problem of Leakage-Minimal Design is a convex optimization problem; we also show that the Karush-Kuhn-Tucker conditions are necessary and sufficient for solving these convex optimizations.

Section IV is about universality. We first show a negative result: that in general, universality breaks down when constraints other than size are added. We can interpret this result as suggesting that universality requires a high degree of symmetry. We then show a positive result about the existence of universally optimal channels for “graph-like” cloaking constraints where “symmetry” may be absent.

We will then explore (Section V) the case when the attacker has uncertainty about the prior, in the sense that he knows

that the prior distribution can be one of a (finite) number of possible distributions on the secret. We show that in this case the problem can be simplified by assuming an “averaged prior”. In particular, universal optimality can be achievable if the constraint is either per the setting in Section IV or the setting in [16]. In particular, all of the previous results hold just by considering a single “averaged prior”.

Finally, in Section VI, we will show an application of these ideas to countermeasures of timing attacks by introducing and analysing “randomized bucketing” strategies.

Related Literature: This paper builds on [16] and makes several important non incremental advances on that work as described in the contributions, both in terms of extending universality results, showing limitations of universality and demonstrating applications of the theory.

More generally this work contributes to the foundations of quantitative security and its results are relevant to most approaches to Quantitative Information Flow, both the ones using Shannon (e.g. [12]), Min Entropy (e.g. [15]), Bayes risk [17] and more recent work using g-leakage [10], [18] and unconditional security [19]. Karush-Kuhn-Tucker conditions for Shannon leakage analysis were used in [20], [21] in a more restricted setting than the one studied here.

II. GENERAL MODEL

Let S represent the *secret* as a random variable. It can take one of the $|\mathcal{S}|$ possibilities from the (finite discrete) set of $\mathcal{S} := \{s_1, \dots, s_{|\mathcal{S}|}\}$ with the (categorical) distribution of $\mathbf{P}_S \in \Delta(\mathcal{S})$. That is, $\mathbf{P}_S \in \{(p_s), s \in \mathcal{S} \mid p_s \geq 0 \forall s \in \mathcal{S}, \sum_{s \in \mathcal{S}} p_s = 1\}$. Without loss of generality, we assume that every secret has a strictly positive probability of realization, i.e., $\text{supp}(\mathbf{P}_S) = \mathcal{S}$.¹

We make the worst-case assumption about the *adversaries*: that they know the true probability distribution of the secret. That is, we take \mathbf{P}_S to be publicly known, hence we will simply refer to \mathbf{P}_S as *the prior*. The defender, observing the (realization of the) secret, produces an observable $o \in \mathcal{O}$.

Let $\Omega \subseteq \mathcal{S} \times \mathcal{O}$ define the permissible observables per each secret. Specifically, if $(s, o) \notin \Omega$, then for secret s , the defender cannot produce observable o . A deterministic channel, denoted by d , is a mapping from secrets to observables, such that each assignment is permissible. Hence, the set of all deterministic channels is: $\{d : \mathcal{S} \rightarrow \mathcal{O} \mid \forall s : (s, d(s)) \in \Omega\}$.

A probabilistic channel, which we denote by δ , allows randomization over permissible observables per each secret. Specifically, the space of probabilistic channels is $\{\delta : \mathcal{S} \rightarrow \Delta(\mathcal{O}) \mid \forall s \in \mathcal{S}, \forall o \in \text{supp}(\delta(s)) : (s, o) \in \Omega\}$. Clearly, any deterministic channel can be represented as a probabilistic channel as well, with degenerate distributions. For the rest of the paper, unless explicitly stated, by “channel” we mean a probabilistic channel.

We use the familiar notation of conditional probability, i.e., $\delta(o|s)$, to designate the probability at which channel δ

¹Support of a probability distribution is defined as the set of all possible values that has a strictly positive probability of realization.

produces observable o when the secret is s . Using this notation, the space of channels is specified by the following conditions:

$$\delta(o|s) \geq 0 \quad \forall (s, o) \in \mathcal{S} \times \mathcal{O} \quad (1a)$$

$$\sum_{o \in \mathcal{O}} \delta(o|s) = 1 \quad \forall s \in \mathcal{S} \quad (1b)$$

$$\delta(o|s) = 0 \quad \forall (s, o) \notin \Omega. \quad (1c)$$

Conditions (1b) and (1c) just impose that the channel should be a legitimate conditional distribution. We will refer to (1c) as “hard” constraints, as they strictly forbid some secret-observable pairs “path-wise”, that is, per each realization of the secret. As a consequence, an adversary can eliminate the forbidden secrets for an observable when making an inference. The naming is to contrast with the “soft” constraint discussed later, which is expressed in terms of an expected value.

The aim is designing channels within the above boundaries that are “leakage-optimal”. That is, finding a feasible conditional distribution $\delta(o|s)$ that, in a quantifiable way, leaks the least information about the secret to an adversary, who is aware of both the distribution of the secret and our channel design.

At a high level, the information “leakage” can be quantified as the difference between the “prior uncertainty” of an adversary and its “posterior uncertainty”, on average. The expected uncertainty is quantified by an *entropy* measure. We will denote the entropy of a random variable \mathcal{S} with probability distribution \mathbf{P} by $\mathcal{H}(\mathcal{S}) = H(\mathbf{P})$, where H is a function from probability distributions to real numbers.

Intriguingly, there are many different candidates for the choice of the entropy, each with a distinct “operational” significance. For instance, the *1-guess-error-probability* is computed as $H(\mathbf{P}) = 1 - \|\mathbf{P}\|_\infty$, where $\|\mathbf{P}\|_\infty$ denotes the ∞ -norm of \mathbf{P} , that is $\|\mathbf{P}\|_\infty = p_{[1]} = \max_i(p_1, \dots, p_n)$. Notably, $1 - \|\mathbf{P}\|_\infty$ is the probability that the best guess of an adversary about the secret is incorrect. A closely related measure is *Min-entropy*: $H(\mathbf{P}) = -\log \|\mathbf{P}\|_\infty$. The *l-guess-error-probability* extends to the cases where an adversary can submit l best guesses. Specifically, $H(\mathbf{P}) = 1 - \sum_{i=1}^l p_{[i]}$ is the probability that none of these guesses would be correct, where $p_{[i]}$ denotes the i th largest element of \mathbf{P} , breaking ties arbitrarily. Another frequently used entropy with a clear interpretation is *guesswork* (*guessing*) entropy: $H(\mathbf{P}) = \sum_{i=1}^n ip_{[i]}$. This represents the minimum expected number of steps that takes a sequentially guessing adversary to get to the secret. Probably the most well-known entropy is the (*Gibbs*)-*Shannon*’s: $H(\mathbf{P}) = -\sum_{i=1}^n p_i \log(p_i)$, pertaining to the shortest coding of the secret, which can also be interpreted as the least expected number of “subset-membership” questions of an adversary before getting to the secret. A family of entropies is known as *Rényi* entropies, parametrised by $\alpha \geq 0$, $\alpha \neq 1$: $H_\alpha(\mathbf{P}) = \frac{1}{1-\alpha} \log(\sum_{i=1}^n p_i^\alpha)$, or equivalently, $H_\alpha(\mathbf{P}) = \frac{\alpha}{1-\alpha} \log \|\mathbf{P}\|_\alpha$, where $\|\cdot\|_\alpha$ denotes the α -norm. Rényi entropies can recover Shannon and Min-entropy as limit cases by respectively letting $\alpha \rightarrow 1$ and $\alpha \rightarrow \infty$. The case of $\alpha = 2$, i.e., $H_2(\mathbf{P}) = -\log \sum_{i=1}^n p_i^2$ is called the *Collision* entropy.

Likewise, $\alpha = 0$ case, i.e., $H_0(\mathbf{P}) = \log |\text{supp}(\mathbf{P})| = \log n$ is known as the *Hartley* entropy.

The uncertainty of the adversary after observing the output of the channel (on average) is measured by *posterior entropy* or “equivocation”, which we denote by $\mathcal{H}[S|O]$. For each of the aforementioned entropies, a posterior entropy can be defined in a meaningful way. For instance, the posterior 1-guess-error-entropy can be simply defined as the average failure rate of an adversary that makes a best guess about the secret after seeing the observable. For a formal representation, first let us define \mathcal{O}^+ to be the set of observables that have a strictly positive probability of realization (given a channel δ). Note that, since $\text{supp}(\mathbf{P}_S) = \mathcal{S}$, we simply have: $\mathcal{O}^+ = \cup_{s \in \mathcal{S}} \text{supp}(\delta(s))$. Using this notation, we can write the following relation for the posterior entropy with respect to 1-guess-error-probability: $\mathcal{H}(S|O) = \sum_{o \in \mathcal{O}^+} p(o) (1 - \|\mathbf{P}_{S|o}\|_\infty)$. Similarly, with respect to guesswork, we can write: $\mathcal{H}(S|O) = \sum_{o \in \mathcal{O}^+} p(o) (\sum_{i=1}^n ip_{[i]}(o))$. For Shannon entropy, we have: $\mathcal{H}(S|O) = \sum_{o \in \mathcal{O}^+} p(o) [-\sum_{s \in \mathcal{S}} p(s|o) \log(p(s|o))]$. For the Rényi family, there are at least two different relations for the posterior entropy in the literature (e.g. [24], [25]):

$$\mathcal{H}[S|O] = \frac{-1}{\alpha-1} \log \left(\sum_{o \in \mathcal{O}^+} p(o) \|\mathbf{P}_{S|o}\|_\alpha^\alpha \right); \quad (2a)$$

$$\mathcal{H}[S|O] = \frac{-\alpha}{\alpha-1} \log \left(\sum_{o \in \mathcal{O}^+} p(o) \|\mathbf{P}_{S|o}\|_\alpha \right). \quad (2b)$$

As in the spirit of [16], we consider a generic conditional entropy that encompass all of the aforementioned entropies. In particular, it has the following structure:

$$\mathcal{H}[S|O] = \eta \left(\sum_{o \in \mathcal{O}^+} p(o) F(\mathbf{P}_{S|o}) \right), \quad (3)$$

where $\eta : \mathbb{R} \rightarrow \mathbb{R}$ function, and F is a bounded scalar function on probability distributions with the following properties:

- *symmetry*, i.e., its value only depends on the shape of a probability distribution and does not change with any re-labelling of the probabilities;
- *expansibility*, i.e., its value does not change by padding the probability distribution with zero entries; and
- *non-decreasing in each element*;

Moreover, one of the following two situations holds:²

$$\eta: \text{increasing, and } F: \text{concave; or} \quad (4a)$$

$$\eta: \text{decreasing, and } F: \text{convex.} \quad (4b)$$

Note that the form of the conditional entropy in (3) governs the form of the unconditional entropy as well (e.g. by taking O and S to be independent). Specifically, $\mathcal{H}[S] = \eta(F(\mathbf{P}))$. For the rest of the paper, unless explicitly clarified, by “entropy” we mean any member of the generic class described above.

Next, we show how this structure can encompass all of the previous entropies as special cases. For instance, Shannon

² $F(\mathbf{P})$ is concave (resp. convex) in \mathbf{P} iff: $\forall \lambda \in [0, 1], \mathbf{P}_1, \mathbf{P}_2 \in \Delta\mathcal{S}$, we have: $\lambda F(\mathbf{P}_1) + (1-\lambda)F(\mathbf{P}_2) \leq$ (resp. \geq) $F(\lambda\mathbf{P}_1 + (1-\lambda)\mathbf{P}_2)$.

entropy can be represented by taking η to be the identity function, i.e., $\eta(x) = x$, and $F(\mathbf{P}) = -\sum_{i=1}^n p_i \log(p_i)$ which is well known to be a symmetric concave function over the space of probability distributions. Likewise, for l -guess-error-probability as well as guesswork, η can be taken as the identity function. For the conditional Rényi entropy as per (2a), we can take $\eta(x) = \frac{-1}{\alpha-1} \log(x)$ on \mathbb{R}^+ and $F(\mathbf{P}) = \|\mathbf{P}\|_\alpha^\alpha = \sum_{i=1}^n p_i^\alpha$. For the conditional Rényi entropy as in (2b), we can take $\eta(x) = \frac{-\alpha}{\alpha-1} \log(x)$ on \mathbb{R}^+ and $F(\mathbf{P}) = \|\mathbf{P}\|_\alpha$. For both cases, F is a symmetric function that is non-decreasing in each element. Moreover, when $0 \leq \alpha < 1$, in both cases, η is increasing and F is concave, and when $\alpha > 1$, η is decreasing and F is convex.

As we argued before, the aim is to design channels that have the lowest leakage of information about the secret while satisfying a set of operational constraints, and the leakage is defined as the difference between the posterior and prior entropies. First point to note is that the choice of the channel cannot change the prior entropy, as the prior entropy of the secret is entirely governed by its prior distribution, which we assume is a “given” parameter that the defender cannot control. Therefore, the problem of minimizing the leakage becomes equivalent to maximizing the posterior entropy (equivocation).

Before we express the general form of the optimal channel design problem, we discuss an additional type of constraint that may be relevant. There are many interesting cases where it may be “feasible” to assign the same observable for all secrets, but such a move may result in a huge deterioration in the system’s quality of the service (QoS). In such cases, the goal is to strike an optimal “balance” between information leakage and QoS. This is for instance the setting in geo-location privacy-utility trade-off [7]–[9] and secrecy-delay trade-off in bucketing as a defence against timing attacks [2], [3].

In its most basic form, the QoS can be captured as an expected value of a “payoff” (desirability) function. In particular, let $u : \mathcal{S} \times \mathcal{O} \rightarrow \mathbb{R}$ where $u(s, o)$ represents how good the realized observable is for a particular secret. Let $\mathbb{E}_{S, O}[U]$ be the expected value of the pay-off, where the expectation is taken with respect to the joint random variable of (S, O) . For conciseness, we drop the subscripts of S, O from the expectation but in order to explicitly show the dependence on the channel, we will use the notation of $\mathbb{E}_\delta[U]$:

$$\mathbb{E}_\delta[U] = \sum_{s \in \mathcal{S}} p_s \sum_{o \in \mathcal{O}} \delta(o|s) u(s, o) \quad (5)$$

$\mathbb{E}_\delta[U]$ can be a metric for the QoS of the channel. The channel design problem then becomes a “two-objective” optimization: (a) minimizing leakage, and (b) maximizing the QoS. The solution concept for multi-objective optimizations is of “Pareto-efficiency” (Pareto-optimality), which are the solutions with a guarantee that no alternative can simultaneously improve all of the objectives (at least one of them strictly). One of the standard methods of converting a multi-objective optimization (MOO) to (a series of) single-objective optimizations (SOO) is to present all but one of the objectives as inequality constraints. Specifically, we can introduce a lower threshold u_{\min} on the

QoS by imposing: $\mathbb{E}_\delta[U] \geq u_{\min}$. Then by varying the value of u_{\min} and solving the resulting single-objective optimizations, the Pareto-frontier (the set of Pareto-optimal solutions) will be found (see e.g. [26]). Hence, with this in mind, for the rest of the paper, we will be dealing with single-objective optimizations. We will refer to the constraint of $\mathbb{E}_\delta[U] \geq u_{\min}$ as the “soft” constraint, since it is expressed in terms of the expected-value, distinguishing it from the “hard” constraints represented by Ω , which are per each realization of the secret.

Putting things together, the optimal channel design problem in its most general form becomes:

$$\max_{\delta \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{O}|}} : \mathcal{H}_\delta[S | O] = \eta \left(\sum_{o \in \mathcal{O}^+} p(o) F(\mathbf{P}_{S|o}) \right), \quad (6)$$

where $p(o) = \sum_{s' \in \mathcal{S}} p_{s'} \delta(o|s')$ and $\mathbf{P}_{S|o}$ is the $|\mathcal{S}|$ -sized conditional probability vector whose entries are $p(s | o) = p_s \delta(o|s) / (\sum_{s' \in \mathcal{S}} p_{s'} \delta(o|s'))$. The constraints of the optimization are as follows:

$$\delta(o|s) \geq 0 \quad \forall o \in \mathcal{O}, s \in \mathcal{S} \quad (7a)$$

$$\sum_{o \in \mathcal{O}} \delta(o|s) = 1 \quad \forall s \in \mathcal{S} \quad (7b)$$

$$\mathbb{E}_\delta[U] \geq u_{\min} \quad (7c)$$

$$\delta(o|s) = 0 \quad \forall (s, o) \notin \Omega \quad (7d)$$

Before we get to our analysis, we present two minimalistic examples to instantiate the constraints. We will return to these examples in Sections IV, as they serve as a counter-example for existence of a universally optimal channel. The first toy example is motivated by geo-location privacy.³ Fig. 1 depicts 4 locations s_1 to s_4 , where the configuration is a representation of their relative positions. The defender is in one of these 4 locations and generates an observable, which can be its reported coordinates, based on which, it receives a location-based service (LBS). Suppose in particular, that s_1 and s_2 are near enough that the same observable can be reported for both of them, but s_1 is too far from s_3 and s_4 such that reporting the same coordinates with them is either infeasible (e.g. it will then not get any network connectivity from an access point) or it will be unacceptable (the quality of the received utility will be too poor). Moreover, s_2, s_3 and s_4 are close enough to produce the same observable. If we label the observables simply by the subset of the secrets that can produce them, then the set of admissible secret-observable pairs, i.e., Ω , is $\{(s_1, \{s_1\}), (s_2, \{s_2\}), (s_3, \{s_3\}), (s_4, \{s_4\}), (s_1, \{s_1, s_2\}), (s_2, \{s_1, s_2\}), (s_2, \{s_2, s_3\}), (s_3, \{s_2, s_3\}), (s_3, \{s_3, s_4\}), (s_4, \{s_3, s_4\}), (s_2, \{s_2, s_3, s_4\}), (s_3, \{s_2, s_3, s_4\}), (s_4, \{s_2, s_3, s_4\})\}$. This Ω determines the hard constraints on the problem, e.g., we must have: $\delta(\{s_2, s_3, s_4\}|s_1) = 0$ because $(s_1, \{s_2, s_3, s_4\}) \notin \Omega$.

As another example, consider a minimalistic bucketing example depicted in Fig. 2. The axis denotes time duration, and s_1 to s_4 represent the distinct execution times of four distinct

³Note that each of these contexts of course have their idiosyncrasies that are abstracted away for the purpose of this paper.

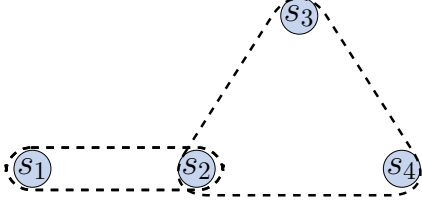


Fig. 1. (Toy example 1) The “secret” is one of the four possible locations s_1 to s_4 . s_1 is located too far away from s_3 and s_4 for all of the secrets to be able to produce the same observable. To avoid clutter, only two of the feasible observables, $\{s_1, s_2\}$ and $\{s_2, s_3, s_4\}$ are demarcated here.

(encryption or decryption) processes, i.e., process 1 takes s_1 time to finish, and so on. If the result of each process is released immediately upon finishing, then they can be uniquely identified just by the timing “side channel”. The result of a finished process can be deferred and released at a later time, to become identical to other processes that take longer to finish. This superset duration time constitutes a *bucket*. In the figure, the arrows represent whether a secret can be deferred till the finishing time of a longer processes. Specifically, suppose that the delay limitation for processes 1 does not allow it to be released as late as s_3 or s_4 . Therefore, the hard constraints can be identically represented as in the previous toy example.

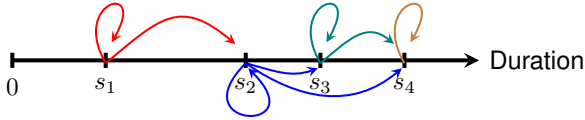


Fig. 2. (Toy example 2) The “secrets” are one of the four processes each with a distinct execution time s_1 to s_4 . The arrows denote which process can be deferred to be released at a later finishing time. For instance, process 2 can be either release instantaneously, i.e., at s_2 , or deferred until s_3 , or until s_4 . In contrast, s_1 cannot be deferred as late as s_3 or s_4 .

Before we present our technical results, we would also like to point out that all of our results naturally extend to the case of generalized gain-based entropies defined in [16] as well:

$$\mathcal{H}_g[S | O] = \eta \left(\sum_{o \in \mathcal{O}^+} p(o) \|GP_{S|o}\|_1 F \left(\frac{GP_{S|o}}{\|GP_{S|o}\|_1} \right) \right)$$

where G is a given $|\mathcal{W}| \times |\mathcal{S}|$ matrix with positive elements where \mathcal{W} is a bounded discrete set, and $\|\cdot\|_1$ is the 1-norm of a vector. When F is the ∞ -norm, the interpretation of the $G(w, s)$ is that it is the gain of the adversary for choosing the guess of $w \in \mathcal{W}$ when the actual secret is $s \in \mathcal{S}$ (this special case is connected to g -leakage as described e.g. in [18]).

III. MINIMAL LEAKAGE IS CONVEX PROGRAMMING

Our first (and the most positive) proposition establishes that for any choice of the entropy and payoff function, the problem of finding an optimal channel is a “convex optimization” (a.k.a “convex programming” [27], [28]). This is a useful result, because convex optimizations have desirable characteristics, e.g., many efficient algorithms for solving them exist (e.g.

interior methods [28]). Moreover, any local optimum has the guarantee to also be a global optimum, so in particular any “descent” algorithm will necessarily converge to a global optimum. Additionally, in Proposition 2, we show that the Karush-Kuhn-Tucker (KKT) conditions fully describe the optimal channel (represent necessary and sufficient conditions of optimality). Later, in Section IV, we demonstrate how this can be used to establish non-trivial properties of the optimal channels. Keep in mind that in our analytical parts, we assume that the (conditional) entropy follows the generic form of (3) where η and F satisfy the conditions described after (3).

Proposition 1: The optimization problem of (6) with constraints (7) for any choice of the pay-off and entropy functions is a convex programming.⁴

Proof: First thing to note is that η from (6) can be simply ignored for both cases, since it is a monotonic $\mathbb{R} \rightarrow \mathbb{R}$ function. Now, suppose we are dealing with case-4a (the argument for case-4b is identical). Our optimization variable is $\delta \in \mathbb{R}^{|\mathcal{S}||\mathcal{O}|}$. In particular, consider it as a $|\mathcal{S}||\mathcal{O}| \times 1$ vector. All we need to show is that: (a) the constraints of the optimization define a convex subset of $\mathbb{R}^{|\mathcal{S}||\mathcal{O}|}$; and (b) the objective function of the maximization is concave in δ .

Establishing (a) is simple: the constraints (7a), (7b) and (7d) trivially define a convex subset. The (7c), as is evident from (5), is also a linear transformation of δ – where the coefficient of $\delta(o|s)$ is $p_s u(s, o)$. Hence, the constraints of the problem define a convex subset of $\mathbb{R}^{|\mathcal{S}||\mathcal{O}|}$. In fact, they define a bounded polyhedron, as the feasible set is the intersection of half-spaces and it does not contain a whole line.

We establish part (b) by writing the objective function as a composition of a series of transformation each of which preserves convexity/concavity. First, note that the expression for the objective function in (6) (ignoring η) in terms of δ is:

$$\sum_{o \in \mathcal{O}^+} p(o) F(P_{S|o}) = \sum_{o \in \mathcal{O}^+} \left[\sum_{s' \in \mathcal{S}} p_{s'} \delta(o|s') \right] F \left(\frac{(p_s \delta(o|s))_{s \in \mathcal{S}}}{\sum_{s' \in \mathcal{S}} p_{s'} \delta(o|s')} \right)$$

where $(p_s \delta(o|s))_{s \in \mathcal{S}}$ represent the $|\mathcal{S}|$ -sized vector, whose entries are $p_s \delta(o|s)$. It is sufficient to show that each term of the (outside) summation is a convex function in δ . Without loss of generality, we show this for $o_1 \in \mathcal{O}^+$:⁵

- Affine transformation 1: $h_1 : \mathbb{R}^{|\mathcal{S}||\mathcal{O}|} \rightarrow \mathbb{R}^{|\mathcal{S}|}$, as $h_1(\vec{y}) = A\vec{y}$, where A is a $|\mathcal{S}| \times |\mathcal{S}||\mathcal{O}|$ matrix whose entries are as follows: $A(i, i) = p_{s_i}$ for $i = 1, \dots, |\mathcal{S}|$, and zero otherwise; Affine transformation is both a convex and a concave function; The result of $A\delta$ is the $|\mathcal{S}| \times 1$ vector of $(p_s \delta(o_1|s))_{s \in \mathcal{S}}$.
- Composition with the concave function F . Recall that by assumption F is non-decreasing in each element and

⁴Both minimizing a convex function and maximizing a concave function, over a convex set, are instances of “convex” programming.

⁵Note that since F is bounded, we have: $\lim_{p(o) \rightarrow 0} p(o) F(P_{S|o}) = 0$, and hence we can only focus on $o \in \mathcal{O}^+$.

concave over probability distributions (case-4a). Composition of a concave function that is non-decreasing in each element with a concave function yields a concave function [27, Page 86].

- Affine transformation 2: consider the transformation $h_2 : \mathbb{R}^{|\mathcal{S}|} \rightarrow \mathbb{R}^{|\mathcal{S}|+1}$ as follows: $h_2(y) = (\vec{y}, \|\vec{y}\|_1)$. Note that $\|\vec{y}\|_1$ is a linear transformation (simply, sum of its elements). The h_2 transformation is equivalent to multiplication by the $(|\mathcal{S}| + 1) \times |\mathcal{S}|$ matrix B , which is composed of the $|\mathcal{S}| \times |\mathcal{S}|$ identity matrix with an extra bottom row of “all ones”.
- Perspective: consider the *perspective* transformation $h_3 : \mathbb{R}^{|\mathcal{S}|+1} \rightarrow \mathbb{R}^{|\mathcal{S}|}$ as follows: $h_3(y, t) = tF(y/t)$. Then, if F is concave, so is h_3 [27, Chapter 3.2.6].

Composition of these together gives us the first term of the expression, establishing the proposition. ■

As mentioned before, a fundamental property of convex optimizations is that any local optimum is a global optimum. In what follows, we establish another important property of the optimal channel design problems: that the Karush-Kuhn-Tucker (KKT) conditions provide both necessary and sufficient conditions for optimality. We will showcase the use of such conditions in Section IV, where we prove existence of universal optimal channels for a special class of constraints. For an overview of the Lagrangian duality and KKT conditions the reader can consult with the rich literature on convex programming such as [27, Ch.5] and [29, Ch.28].

Proposition 2: KKT conditions are necessary and sufficient for finding the optimal channel design, described by (6), (7).

Proof: We start by noticing that in the most basic form, KKT conditions are expressed for cases where the function in the objective and constraints are “continuously differentiable”, whereas some of our convex objective functions (e.g. in the case of min-entropy or guesswork) are piecewise linear. There is however a simple and standard translation from piecewise-linear convex functions into continuously differentiable functions by forming the epigraph problem [27, §5.2.5].

The proof is straightforward: all of our constraints (7) are affine hence the KKT conditions are necessary – this is known as “Linearity Constraint Qualification” (LCQ). Moreover, since we showed that these problems are convex optimizations, the KKT conditions are also sufficient [27, §5.5.3]. ■

The “Lagrangian” for the problem of (6) with constraints (7), denote by L is:

$$L = \sum_o \left[\sum_{s'} p_{s'} \delta(o|s') \right] F \left(\frac{(p_s \delta(o|s))_{s \in \mathcal{S}}}{\sum_{s'} p_{s'} \delta(o|s')} \right) + \sum_{s,o} \lambda_o^s \delta(o|s) + \sum_s \mu_s \left(\sum_o \delta(o|s) - 1 \right) + \rho \left(\sum_{s,o} p_s \delta(o|s) u(s, o) - u_{\min} \right) + \sum_{(s,o) \notin \Omega} \gamma_o^s \delta(o|s) \quad (8)$$

where the multipliers μ, γ are from the equality constraints (7b) and (7d), and are therefore free (no sign constraint), whereas the multipliers λ, ρ pertain to inequalities (7a) and (7c), and are hence required to be positive (dual feasibility).

The optimization problem then becomes equivalent to solving the following KKT conditions:

- 1) Vanishing first order derivatives of L with respect to each of the optimization variables $\delta(o|s)$, that is, $\nabla L = \vec{0}$ (where ∇ is the gradient with respect to the (primal) variables $\delta(o|s)$). That is, for each $\delta(o|s)$: $\frac{\partial L}{\partial \delta(o|s)} = 0$;
- 2) Primal feasibility: constraints (7a)–(7d);
- 3) Dual feasibility: $\lambda_o^s \geq 0$, $\forall s, o$, and $\rho \geq 0$;
- 4) Complementary slackness: $\forall s, o$ $\lambda_o^s \delta(o|s) = 0$ and $\rho (\sum_{s,o} p_s \delta(o|s) u(s, o) - u_{\min}) = 0$.

IV. UNIVERSALITY (MEASURE-INVARIANCE) RESULTS

Although Proposition 1 holds for any choice of the entropy and utility, it is a much weaker statement than the notion of *universality* as measure-invariance optimality described in [16]. In particular, they showed that for a special choice of the constraints, there is a “universally optimal” randomized channel, in the sense that, there exists a channel that is leakage-optimal irrespective of the choice of the entropy. That is, all of the optimizations share a common optimizer.

A. Universal (Measure-Invariant) Optimality

Here, we give an overview of the main results on universal (measure-invariant) optimality proved in [16]. Informally, the result says that if the only constraint on the channel is that at most $k < |\mathcal{S}|$ of secrets can produce the same observable, then it is possible to design a channel which is universally optimal. The design constraint can be expressed as follows: the size of the pre-image of any observable must be at most k . To design such channel one starts by sorting the prior in descending order, and identifying an index j^* , which is the index of the first “non-giant” secret, that is, the last secret with a “too large” probability compared to the remaining secrets. This way, the first $j^* - 1$ secrets will constitute the “giants”.

More formally given a prior $\mathbf{P} = (p_1, \dots, p_n)$, sorted in descending order and an integer $k < n$, let index j^* be:

$$j^* := \min \left\{ j : 1 \leq j \leq k, p_j \leq \frac{\sum_{i=j}^n p_i}{k - j + 1} \right\}. \quad (9)$$

With j^* defined as above, let $\vec{\pi}$ denote the following probability distribution over k elements:

$$\vec{\pi} := \left(p_1, \dots, p_{j^*-1}, \frac{\sum_{i=j^*}^n p_i}{k - j^* + 1}, \dots, \frac{\sum_{i=j^*}^n p_i}{k - j^* + 1} \right) \quad (10)$$

i.e. $\vec{\pi}$ is a k -sized probability distribution whose first $j^* - 1$ probabilities are the first $j^* - 1$ probabilities of the prior (the “giants”), and the remaining probabilities are mashed together and spread “uniformly”. The universal optimality result in [16] (re-phrased) is the following:

Theorem 1: Consider channels form secrets with a given prior $\mathbf{P} = (p_1, \dots, p_n)$ to a set of observables, that satisfy the constraint that at most k secrets can produce the same observable, i.e., that the size of the pre-image of any observable is at most k , where $k < n$. Let $\vec{\pi}$ be defined as in (10). Then the maximum posterior entropy achieved by any such channel is $H(\vec{\pi})$ for any Schur-concave choice of entropy function H .

Moreover, if the only design constraint is the pre-image size constraint, then there exists a channel that achieves a posterior entropy exactly equal to $H(\bar{\pi})$ for any Schur-concave H , and is hence universally (measure-invariantly) optimal.

Notice that $\bar{\pi}$ is independent from the choice of the entropy H . Recall that since the prior distribution cannot be changed in this setting, the posterior-entropy maximizing channel in the theorem is also (measure-invariantly) leakage-optimal. The proof in [16] is constructive, in that it provides the optimal channel explicitly. The construction is non-trivial and is left out for brevity. We just mention that the optimal channel is constructed such as to guarantee that for any observable, the posterior (Bayesian) probability over the secrets is exactly $\bar{\pi}$.

For both of our toy examples in Figures 1 and 2, only the first part of the theorem can be applied, i.e., the upper-bound of the posterior entropy with $k = 3$. The second part of the Theorem does not apply, because the design constraints in nether of the toy examples can be reduced to just a pre-image size constraint. Specifically, in both examples, there is only one allowed pre-image of size 3, while all subset of size 3 of the four secrets should be allowed as a pre-image. Thus, existence of a measure-invariant channel is not guaranteed.

The positive results in [16] raise a fundamental question: does this strong notion of universality hold in general? Next, as one of the main contributions of this paper, we settle this question through a counter-example.

Proposition 3: In general, there is no universally optimal channel, i.e., for a given prior and a set of design constraints, there is no channel that minimizes the leakage for any choice of how the leakage is measured.

In other words, in general, the problem of designing a leakage minimal channel is sensitive to the choice of entropy, i.e., the way leakage is quantified.

Proof: Consider the following example: There are 4 secrets: $\mathcal{S} = \{1, 2, 3, 4\}$ with prior $\mathbf{P} = (p_1, p_2, p_3, p_4)$. The set of observables (outputs) is $\{a, b\}$. The set of feasible observables is defined by $\Omega = \{(1, a), (2, a), (2, b), (3, b), (4, b)\}$. That is, for secret 1, the only possible observable to show is a , for secret 2, both a and b are allowed, and for secrets 3 and 4, the only allowed observable is b . There is no soft utility constraint. Following the admissible observables for secrets 1, 3 and 4, we have: $\delta(b|1) = \delta(a|3) = \delta(a|4) = 0$, and therefore: $\delta(a|1) = \delta(b|3) = \delta(b|4) = 1$. For secret 2, $\delta(a|2)$ and $\delta(b|2)$ are free, as long as they are positive and add up to 1. Therefore, $\delta(a|2)$ is the only variable of optimization. We will therefore introduce the variable x defined as $x := p_2\delta(b|2)$.

The probability that a is observed is: $p(a) = \sum_{i=1}^4 p_i\delta(a|i)$, which, following the previous argument, will reduce to: $p_1 + p_2 - x$. Similarly, $p(b) = x + p_3 + p_4$. Hence, the problem of maximizing posterior entropy reduces to the following single-variable optimization:

$$\begin{aligned} & \text{Maximize: } (p_1 + p_2 - x)F\left(\frac{p_1}{p_1 + p_2 - x}, \frac{p_2 - x}{p_1 + p_2 - x}\right) \\ & (x + p_3 + p_4)F\left(\frac{x}{x + p_3 + p_4}, \frac{p_3}{x + p_3 + p_4}, \frac{p_4}{x + p_3 + p_4}\right) \\ & \text{subject to: } 0 \leq x \leq p_2 \end{aligned}$$

For any (differentiable) choice of F , in the light of Proposition 1, we can find the optimal solution by simply taking the derivative of the objective function with respect to x and equating it with zero. For the choice of F as per Shannon, the optimizer is derived as: $x_{\text{Sh}} = \frac{p_2(1 - p_1 - p_2)}{1 - p_2}$. And for Rényi family with $\alpha = 2$ - i.e., the collision entropy and the form of the Rényi conditional entropy as in (2a), it is:

$$\begin{aligned} x_{\text{R}2} &= \frac{\#}{p_3(p_1 + p_2 - 1) + (p_2 - 1)(2p_1 + p_2 - 1) + p_3^2} + p_2 \\ \# &= p_1(p_1(2p_2 + p_3 - 1) + (p_2 - 1)p_3 + (p_2 - 1)^2 + p_3^2) \\ &\quad - \sqrt{p_1^2(p_3(p_1 + p_2 - 1) + (p_1 + p_2 - 1)^2 + p_3^2)} \end{aligned}$$

These two evaluate to different values for instance for the prior of $(0.2, 0.5, 0.15, 0.15)$. Specifically, $x_{\text{Sh}} = 0.3$ and $x_{\text{R}2} = 0.265$ (up to 3 digits). Figure 3 depicts the objective function (posterior entropy) v.s. x for these two and some other choices of the entropy function. The optimizer clearly varies with the choice of F . ■

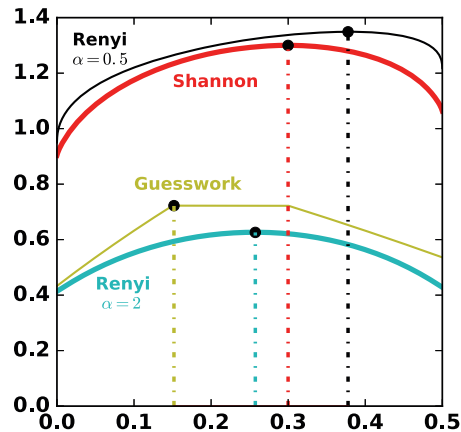


Fig. 3. Posterior entropy v.s. x for different choices of the entropy function for the negative counter-example. The maximizer is for guesswork entropy 0.1518, for Rényi with $\alpha = 2$ is 0.2573, for Shannon entropy is 0.2998 and for half-norm is 0.37493

The constraint in [16] that lead to universal optimality was a cap on the size of the pre-images of the observables. This constraints allows for a level of symmetric flexibility, e.g., any pair of secrets can be conflated with each other, something that our counter-examples do not allow. This leads to another basic question: is the setting in [16] necessary for existence of a universal solution? In what follows, we describe a class of problems beyond the pre-image size-constraint of [16] and establish that they still admit a universally optimal channel.

Consider an instance of our optimal channel design problem where the design constraints are expressed as a (non-directional) graph, where the nodes of the graph are the secrets, and the edges represent the observables. The two ends of an edge determine the two secrets that can produce that

observable. One can also assume that each node has a “self-loop” as well, in that, for each secret, there is always a choice of producing a fully revealing observable. This setting means that only (specific) “pairs” of secrets can conflate with each other. Note that this is *not* a case of [16] with a pre-image size of 2, since, unlike there, not all pairs of secrets are permissible. For instance, if there are 4 secrets $\{1, 2, 3, 4\}$, the setting in [16] with cap-size 2 required that all edges $1 \leftrightarrow 2$, $1 \leftrightarrow 3$, $1 \leftrightarrow 4$, $2 \leftrightarrow 3$, $2 \leftrightarrow 4$, and $3 \leftrightarrow 4$ (besides the self-loops) should be present, in contrast, we can now have an *arbitrary* subset of these be allowed, for example $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, $2 \leftrightarrow 4$, and $3 \leftrightarrow 4$ (besides the self-loops). Critically, “symmetry” is now broken. We will refer to this class of optimal channel design problems, where there is no soft constraint, and the hard constraints are expressible by such a graph, as *graph-constrained*. Note that in the absence of a soft constraint, and the assumption that a fully-revealing observable for each secret is always possible, the problem is always “feasible”.

In our general optimal channel design problem in Section II, we designated the set of admissible secret-observable pairs by Ω . Let us translate the above graph-based representation of the constraints with respect to Ω . For an observable $o \in \mathcal{O}$, let the notation o^{-1} represent the set of secrets that are allowed to produce o , that is, $o^{-1} := \{s \in \mathcal{S}, (s, o) \in \Omega\}$. Then our positive result applies to any case where $\forall o \in \mathcal{O}$, we have: $|o^{-1}| \leq 2$.

Proposition 4: Any “graph-constrained” optimal channel design problem admits a universally optimal solution.

Proof: We present the proof in the following steps: First, we take the entropy to be the special case of Shannon, and investigate its optimal solution (whose existence is guaranteed). In particular, we investigate its KKT conditions as “necessary” conditions of optimality. Next, we show that these KKT conditions are sufficient to satisfy the KKT conditions of any other entropy as well. By the “sufficiency” of KKT, this establishes that a Shannon-optimal channel is also optimal for any choice of entropy as well. Keep in mind that in Proposition 2, we showed that the KKT conditions are both necessary and sufficient for optimality of a channel.

Let us start with the first-order conditions of KKT. Taking the partial derivative of the Lagrangian in (8) with respect to $\delta(o_1|s_1)$ for a $(s_1, o_1) \in \Omega$, after removing the cancelling terms, and equating the result to zero, yields the following:

$$-p_{s_1} \log\left(\frac{p_{s_1} \delta(o_1|s_1)}{\sum_{s'} p_{s'} \delta(o_1|s')}\right) + \mu_{s_1} + \lambda_{o_1}^{s_1} = 0 \quad (11)$$

which can be written simply as $-p_{s_1} \log(p(s_1|o_1)) + \mu_{s_1} + \lambda_{o_1}^{s_1} = 0$. Also, the complementary slackness condition states that $\lambda_o^s \delta(o|s) = 0 \forall s \in \mathcal{S}, o \in \mathcal{O}$. This in turn implies that, if for a $(s_1, o_1) \in \Omega$, we have $\delta(o_1|s_1) > 0$, i.e., the channel assigns a strictly positive probability to observable o_1 when the secret is s_1 , then we must necessarily have $\lambda_{o_1}^{s_1} = 0$. Therefore, for such (s_1, o_1) , the first order condition further simplifies to:

$$-p_{s_1} \log(p(s_1|o_1)) + \mu_{s_1} = 0$$

In particular, if there are $o_1, o_2 \in \mathcal{O}$ for which both $\delta(o_1|s_1) > 0$ and $\delta(o_2|s_1) > 0$ (same s_1), we must have: $\log(p(s_1|o_1)) = \log(p(s_1|o_2)) = \mu_{s_1}/p_{s_1}$. Since $\log(x)$, $x > 0$, is a strictly increasing function, the above equality is satisfied only when:

$$p(s_1|o_1) = p(s_1|o_2).$$

and since the support of the posterior entropies is at most 2, this further implies that $\mathbf{P}_{S|o_1} = \mathbf{P}_{S|o_2}$.

Now, consider any other entropy. That is, in the Lagrangian (8), F is a generic concave symmetric function. Keep the same optimal channel δ , and the same Lagrange multipliers λ_o^s and γ_o^s ($s \in \mathcal{S}, o \in \mathcal{O}$) as per Shannon’s (that is, take all of the Lagrange multipliers except for μ_s , $s \in \mathcal{S}$) but potentially different values for μ_s (which we designate by μ'_s). The primal feasibility is still satisfied for the new entropy, as the choice of entropy does not affect feasibility of the channel. The dual feasibility as well as complementary slackness constraints are also satisfied, as the same λ_o^s and $\delta(o|s)$ are carried forward. Hence, the only condition that we need to investigate is the vanishing of the first-order derivatives. In particular, we need to see whether whenever $\delta_{o_1}^{s_1} > 0$, we can have: $\partial L / \partial \delta_{o_1}^{s_1} + \mu'_{s_1} = 0$. Note that:

$$\frac{\partial L}{\partial \delta_{o_1}^{s_1}} = p(s_1) \left[F(\mathbf{P}_{S|o_1}) + F_1(\mathbf{P}_{S|o_1})(1 - p(s_1|o_1)) - \sum_{i \neq 1} F_i(\mathbf{P}_{S|o_1}) p(s_i|o_1) \right]$$

where F_i is the partial derivative of F with respect to its i ’th element. If we define $\varphi(\mathbf{P}_{S|o_1}) := F(\mathbf{P}_{S|o_1}) + F_1(\mathbf{P}_{S|o_1})(1 - p(s_1|o_1)) - \sum_{i \neq 1} F_i(\mathbf{P}_{S|o_1}) p(s_i|o_1)$, then the first order condition can be simply written as $p(s_1)\varphi(\mathbf{P}_{S|o_1}) + \mu'_{s_1} = 0$. Recall that there is no constraint (e.g. dual feasibility or complementary slackness) on μ'_s (it is a “free” variable). Taking $\mu'_{s_1} = -p(s_1)\varphi(\mathbf{P}_{S|o_1})$ will satisfy the first-order-condition for F for channel δ , since δ satisfies $\mathbf{P}_{S|o_1} = \mathbf{P}_{S|o_2}$ for all o_2 such that $\delta(o_2|s_1) > 0$, and hence, $p(s_1)\varphi(\mathbf{P}_{S|o_1}) = p(s_1)\varphi(\mathbf{P}_{S|o_2})$, and therefore, the same μ'_{s_1} will satisfy the first order condition with respect to $\delta(o_2|s_1)$ as well. Putting things together, we have found Lagrange multipliers that satisfy the KKT conditions corresponding to F , for the optimal channel corresponding to Shannon, hence, the optimal channel with respect to Shannon is also optimal for F as well (thanks to “sufficiency” of KKT). ■

V. UNCERTAINTY ABOUT THE PRIOR

So far, we assumed the adversary knows the exact prior distribution of the secret. But this may be an unrealistic assumption for some settings. Here, we analyze the setting where the adversary does not know the exact distribution of the prior, but knows that the prior distribution can be one of a number of possibilities, each happening with a known probability (a distribution over distributions⁶).

⁶The adversary’s uncertainty should not be seen here as a subjective belief but as a genuine reflection about the possible state of the system.

At a high level, the main result of this section is the following: the best strategy for the defender is *not* to “customize” its strategy with respect to the context depending on the particular prior given each context, but instead, to build an “averaged prior”, and design the best strategy over this averaged prior and play it irrespective of the contexts. In particular, whenever the constraints are “full cloaking” constraints as in [16], then there exists a universally optimal strategy $\bar{\delta}$.

This result may not be immediately intuitive, as there can be a counter-argument as follows: Among the available priors (conditional probability of the secret given the contexts), there are some particularly “good” ones, in the sense that they are very conducive to hide the secret (e.g. they are very close to uniform in a symmetric constraint setting). Then shouldn’t we adopt the optimal channel for such priors in those contexts, specially, if they have a high probability weight of occurrence? Our result refutes this intuitive argument.

To formalize the setting: the adversary’s uncertainty can be modelled by introducing the “hidden” (discrete) random variable for the *context*, C , that is jointly distributed with the secret. The space of the context is $\mathcal{C} = \{c_1, \dots, c_{|\mathcal{C}|}\}$. Without loss of generality, we assume that the context has full support. The channel designer (the defender), knows the true distribution of the secret. Technically speaking, it “observes” the realization of the context. The adversary, on the other hand, does not directly observe the context, but knows the probability of the realization of each context, P_C , as well as the (conditional) probability distribution of the secret given each context, $P_{S|C}$. Note that knowledge of P_C and $P_{S|C}$ is equivalent to the knowledge of the “joint” probability distribution of the context and the secret $P_{S,C}$.

The adversary only sees the observables and wants to “infer” about the underlying secret. As before, let O denote the (discrete) random variable representing the observable (output). In worst case, one can assume that the adversary knows $P_{O|S,C}$, and hence, using his knowledge of $P_{S,C}$ can use the Bayes’ rule to update his best belief about the secret after observing O , i.e., constructing his *posterior*:

$$p(s | o) = \frac{p(s, o)}{p(o)} = \frac{\sum_{c \in \mathcal{C}} p(c) p(s|c) p(o|s, c)}{\sum_{s' \in \mathcal{S}} \sum_{c \in \mathcal{C}} p(c) p(s'|c) p(o|s', c)}$$

Note that the defender is not directly interested in not leaking information about the context and only cares about the secret, but should be wary of how the adversary can use his information about the joint distribution of the context and secret to intuit about the secret based on the observation. Also, for clarity, we repeat that the adversary does not “observe” the context nor the secret.⁷

The defender decides what observable to produce per each secret in each context, potentially using randomization and benefit from the ambiguity that it can inject. As before, the strategy has to satisfy some operational constraints. We may have hard constraints prescribing which secrets can produce

⁷For the scenario where the adversary can directly observe the context, the problem will reduce to designing $|\mathcal{C}|$ optimal channels according to optimizations as in (6) and (7) with priors $P_{S|c}$ for each $c \in \mathcal{C}$.

which observables, which in part determine which subsets of secrets can be conflated with each other. In the previous sections, we expressed these “hard” operational constraints through $\Omega \subseteq \mathcal{S} \times \mathcal{O}$, representing the set of permissible secret-observable pairs. In the presence of contexts, in the most general form, the permissible observables for a secret may depend on the context as well, and hence Ω should be now a subset of $\mathcal{S} \times \mathcal{C} \times \mathcal{O}$. However, for the result of this section, we assume that these constraints are context-independent, i.e., the same subset of observables is permissible for a secret irrespective of the context, and hence, we keep Ω to be a subset of $\mathcal{S} \times \mathcal{O}$.

Likewise, there can be soft operational constraints in the form of satisfying a minimum expected utility, as in (7c). The expectation is now taken with respect to the context as well, that is, we must have: $\mathbb{E}_{S,C,O}[U] \geq u_{\min}$. However, for the result of this section, we assume that the payoff function, i.e., the measure of “goodness” of each observable for each secret, does not depend on the context. Hence (compare with (5)):

$$\mathbb{E}_{\delta}[U] = \sum_{s,c,o} p(s, c) \delta(o|s, c) u(s, o)$$

As before, without loss of generality, assume that we are dealing with case (4a) where F is concave and η is increasing. Also note that, again, the choice of the strategy cannot affect the prior entropy of the secret. Hence, the problem of designing for minimum leakage is again equivalent to maximizing the posterior entropy. Ignoring η , since it is just an increasing scalar function, the posterior maximization objective in (6) can hence be written as:

$$\begin{aligned} & \sum_o p(o) F(P_{S|o}) \\ &= \sum_o \left(\sum_{s,c} p(s, c) \delta(o|s, c) \right) F \left(\frac{\sum_c \left(p(s, c) \delta(o|s, c) \right)_{s \in \mathcal{S}}}{\sum_{s,c} p(s, c) \delta(o|s, c)} \right) \end{aligned} \quad (12)$$

where by $(p(s, c) \delta(o|s, c))_{s \in \mathcal{S}}$ we mean the $|\mathcal{S}|$ -sized vector whose entry for $s \in \mathcal{S}$ is $p(s, c) \delta(o|s, c)$. The constraint of the optimization are (compare with (7)):

$$\delta(o|s, c) \geq 0 \quad \forall o \in \mathcal{O}, (s, c) \in \mathcal{S} \times \mathcal{C} \quad (13a)$$

$$\sum_{o \in \mathcal{O}} \delta(o|s, c) = 1 \quad \forall (s, c) \in \mathcal{S} \times \mathcal{C} \quad (13b)$$

$$\mathbb{E}_{S,C,O}[U] \geq u_{\min} \quad (13c)$$

$$\delta(o|s, c) = 0 \quad \forall (s, o) \notin \Omega \quad (13d)$$

Given any “context-dependent” strategy δ , we define a corresponding “context-independent” strategy $\bar{\delta}$ as follows:

$$\bar{\delta}(o|s) = \sum_c p(c|s) \delta(o|s, c) \quad (14)$$

To be precise, the strategy is $\bar{\delta}$ such that for any $c' \in \mathcal{C}$, $\bar{\delta}(o|s, c') = \bar{\delta}(o|s)$, i.e., $\bar{\delta}$ represents playing the same randomized strategy of $\bar{\delta}$ irrespective of the context. This context-free strategy is a mixing of the context-dependent strategies

with weights equal to conditional probability of the context given the secret. In other words, $\tilde{\delta}$ “marginalizes away” the dependence of δ on the context.⁸

First, we show that $\tilde{\delta}$ is itself a legitimate strategy:

- 1) $\tilde{\delta}(o|s, c) \geq 0$: trivially (product of non negative terms);
- 2) $\forall (s, c) \in \mathcal{S} \times \mathcal{C} : \sum_{o \in \mathcal{O}} \tilde{\delta}(o|s, c) = 1$; This is because:

$$\begin{aligned} \sum_o \tilde{\delta}(o|s, c) &= \sum_o \sum_{c'} p(c'|s) \delta(o|s, c') \\ &= \sum_{c'} p(c'|s) \sum_o \delta(o|s, c') = \sum_{c'} p(c'|s) = 1 \end{aligned}$$

where we first exchanged the order of the summations, and then respectively used the facts that $\delta(o|s, c')$ and $p(c'|s)$ are conditional distributions.

- 3) We show that $\mathbb{E}_{\tilde{\delta}}[U] = \mathbb{E}_{\delta}[U]$, and hence $\mathbb{E}_{\delta}[U] \geq u_{\min}$ would imply $\mathbb{E}_{\tilde{\delta}}[U] \geq u_{\min}$. For this purpose, we establish the following lemma, which we will use later:

Lemma 1: δ and $\tilde{\delta}$ induce the same (joint) distribution on (S, O) . That is: $p_{\delta}(s, o) = p_{\tilde{\delta}}(s, o) \forall s \in \mathcal{S}, \forall o \in \mathcal{O}$.

Proof: We have:

$$\begin{aligned} p_{\delta}(s, o) &= \sum_c p(c|s) \delta(o|s, c) = \tilde{\delta}(o|s) \\ &= \tilde{\delta}(o|s) \sum_c p(c|s) = \sum_c p(c|s) \tilde{\delta}(o|s, c) = p_{\tilde{\delta}}(s, o) \end{aligned}$$

Now, $\mathbb{E}_{\tilde{\delta}}[U] = \mathbb{E}_{\delta}[U]$ follows as a simple corollary:

$$\begin{aligned} \mathbb{E}_{\tilde{\delta}}(U) &= \mathbb{E}_{\tilde{\delta}}(\mathbb{E}_{\tilde{\delta}}(U|S, O)) = \mathbb{E}_{\tilde{\delta}}(\mathbb{E}_{\tilde{\delta}}(U|S, O)) \\ &= \mathbb{E}_{\tilde{\delta}}(\mathbb{E}_{\delta}(U|S, O)) = \mathbb{E}_{\delta}(U) \end{aligned}$$

The second equality holds because U is invariant with respect to C , and the third equality is due to Lemma 1. The first and last equality is simply the total expectation.

- 4) $\tilde{\delta}(o|s, c) = 0 \forall (o, s) \notin \Omega_s$, trivially. Note that we the assumption that the cloaking constraints do not depend on the context, and only on the secret.

Next, we show that replacing any context dependent strategy with its context-independent would lead to same leakage (irrespective of the choice of the entropy).

Proposition 5: For any given strategy δ , we have: $\mathcal{H}_{\delta}(S|O) = \mathcal{H}_{\tilde{\delta}}(S|O)$.

Proof: This is a direct consequence of Lemma 1, once we notice that $\mathcal{H}(S|O)$ is totally determined by $P_{S,O}$. ■

Our main result of this section:

Proposition 6: The optimization in (12) subject to constraints in (13) can be simplified to the following:

$$\text{Maximize: } \eta \left(\sum_{o \in \mathcal{O}^+} p(o) F(P_{S|o}) \right), \tilde{\delta} \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{O}|} \quad (15)$$

⁸Note that we cannot marginalize away the dependence on the secret because of the secret-dependent constraints. These secret dependent constraints are exactly why the trivial solutions like $\delta(o|s, c) = \text{cte.}$ are not acceptable.

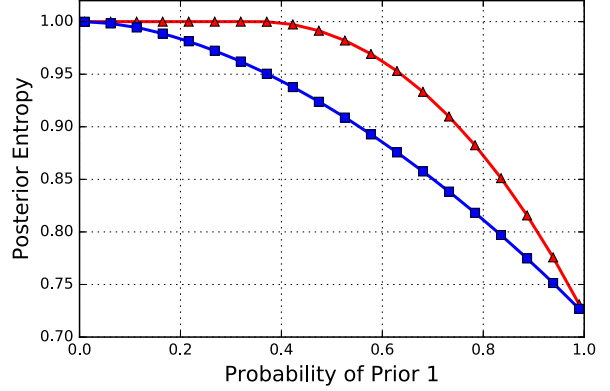


Fig. 4. Comparing the Shannon’s posterior entropy between the optimal design as per Proposition 6 and the heuristic best alternative, where the best channel for each prior is designed and played according to the context. The priors are: $P_1 = (1/3, 1/3, 1/3)$ (the “good” prior) and $P_2 = (0.8, 0.15, 0.05)$ (the “bad” prior). The x-axis is the probability (weight) of P_1 . As we can see, except trivially for the two end-points, the optimal strictly outperforms this “best” heuristic. The cloak size is 2.

where $p(o) = \sum_s p(s) \tilde{\delta}(o|s)$ and $p(s|o) = p(s) \tilde{\delta}(o|s) / (\sum_{s'} p(s') \tilde{\delta}(o|s'))$. The new constraints are:

$$\tilde{\delta}(o|s) \geq 0 \quad \forall o \in \mathcal{O}, s \in \mathcal{S} \quad (16a)$$

$$\sum_{o \in \mathcal{O}} \tilde{\delta}(o|s) = 1 \quad \forall s \in \mathcal{S} \quad (16b)$$

$$\mathbb{E}_{\tilde{\delta}}[U] \geq u_{\min} \quad (16c)$$

$$\tilde{\delta}(o|s) = 0 \quad \forall (s, o) \notin \Omega \quad (16d)$$

Notice that by proposition 6 the optimization problem over a set of priors reduces to an optimization over a single prior, hence whenever the constraints are “full cloaking” constraints then Algorithm 1 from [16] provides a universal optimizer for this uncertainty setting.

Discussion: As we mentioned in the beginning of this section, an alternative heuristic is to play the best channel per each context. One can argue that if the “good” priors that lead to a particularly strong channel have a high probability, it may be better to play this heuristic. However, as we established in proposition 6, this heuristic is wrong. For a numerical depiction, in Fig 4, we have plotted the posterior entropy that is achieved by the optimal strategy $\tilde{\delta}$ per proposition 6 against this heuristic strategy of playing the best channel per each prior. As we can see, for any weight of the two priors (except trivially when the weight is either 0 or 1 where the two strategies become the same), the $\tilde{\delta}$ strictly outperforms the heuristic strategy.

VI. CASE STUDY: OPTIMAL DEFENCE AGAINST TIMING LEAKAGE

As we mention in the introduction, there are many settings in which, an adversary can gain unintended information just by observing the execution time of a process. This is e.g. one

of the side-channel sources considered in [30], [31] where they show secrets like the illnesses, medications or surgeries of the user in healthcare context, her family income and investments in the context of taxation can be revealed by analysing encrypted traffic despite the HTTPS protection. This section explores an application of our framework in the context of defence against timing as side channel leakage.

The timing leakage can be completely eliminated by delaying release of computation results to the maximal possible computation time for that set of possible computations. That is, if all computations take the same time (as the maximum possible time), then the attacker cannot make any time inference on which computation took place. Delaying all computations to this extreme upper bound may however be practically undesirable, as the performance degradation can be too high.

To combat this delay overhead, inspired by the “bucketing” scheme in cryptography [2], [3], [22], [23], several intermediate epochs between the minimum and the maximum computation time can be considered. Each of these intermediate epochs will represent a “bucket”. Each computation can be then delayed up to a later epoch before being released. A simple argument can be made that the release epochs (buckets) should be chosen among the computation times, as there is no point in having a release epoch that falls in between two consecutive computation times (the slower ones cannot be part of the bucket, and the faster ones should not be delayed pointlessly). However, the choice of the number of buckets, their epochs, and process-bucket assignments are non-trivial.

Trade-offs between number of buckets and performance for a cryptographic system with deterministic timing behaviour have been formally investigated by Köpf and Dürmuth in [2]. In particular, they show that in the context of RSA, using five buckets, it is still possible to have a minimal penalty in performance (less than 1% time delay) while significantly weakening timing-channel information leakage.

Here, we consider a different (simpler) setup: while in [2] the secret is the cryptographic key which has correlation with the processing time, we consider the (un-delayed) finishing time itself to be the secret. In other words, we would like to hide from the adversary which process has taken place. We formulate our “randomized bucketing” to give the optimal trade-off between leakage and delay. We compare our randomized bucketing with a heuristic scheme, which we call “deterministic bucketing”, where each process is delayed up to its nearest available release time (bucket), where the number of buckets and their epochs are optimized. Note that our “randomized bucketing” offers more degrees of freedom compared to “deterministic bucketing” in that the release times of a process can be any of the future epochs according to the probability distribution that we design. In particular, any deterministic bucketing scheme is a special case of randomized bucketing too: one in which the entire probability is put on the next immediate available epoch.

In the formal analysis below, leakage will be measured in term of min-entropy, i.e. in terms of the probability of correctly guessing the secret in one try. Both bucketing strategies will

be analysed as optimization problems, in particular as multi-objective optimizations aiming at minimizing both leakage and delay overhead. The problem is set in general terms, in particular the optimal number of buckets and their locations are determined by the optimization itself.

Let us denote the probability that the bucketing strategy sends secret i to bucket j by $\delta(j|i)$. We have given our randomized bucketing as a Linear-Programming in Fig 5. The objective is to minimize:

$$\sum_{j=1}^N p_j \max_{i \leq j \leq N} p_i \delta(j|i) / p_j = \sum_{j=1}^N \max_{i \leq j \leq N} p_i \delta(j|i)$$

(this is expressed by combining $\sum_{j=1}^N z_j$ and $z_j \geq p_i \delta(j|i)$, $i \leq j$, $j \leq N$ in Fig 5). For each bucket j , $\max_{i \leq j \leq N} p_i \delta(j|i) / p_j$ is the highest probability of guessing the secret under δ , hence minimizing $\sum_{j=1}^N \max_{i \leq j \leq N} p_i \delta(j|i)$ is minimizing the expected probability of guessing the secret (in one guess) given the strategy δ . The second objective, i.e. for δ to minimize delay, is given by the constraint $\sum_{j=1}^N \sum_{i=1}^j p_i \delta(j|i) (T_j - T_i) \leq \epsilon$. The term $\sum_{j=1}^N \sum_{i=1}^j p_i \delta(j|i) (T_j - T_i)$ is the expected delay, summed over each bucket (inner summation) and over all buckets (outer summation). This expected delay is imposed to be below ϵ . As ϵ varies the optimization solutions will build the Pareto front.

Our implementation of the optimization of the deterministic bucketing is presented in Fig 6 as a linear integer programming. Here, as the strategy is deterministic, it can only take values 0 or 1, hence the constraints $\delta(j|i) \in \{0, 1\}$ for $i \leq j \leq N$. The constraint that each secret is mapped to its closest bucket is enforced by $\delta(j|k) \geq \delta(j|i)$, $i \leq k \leq j \leq N$. To see this, suppose $\delta(j|i) = 1$, i.e., the strategy assigns i to bucket j . Then the constraint requires any secret k between i and j to satisfy $\delta(j|k) = 1$, i.e., k must be assigned to bucket j too. Finally the expression of the minimum delay constraint is the same as in randomized bucketing.

As we mentioned, a deterministic bucketing strategy is also a randomized bucketing strategy, hence a solution to the randomized bucketing optimization will always outperform a solution of the standard bucketing optimization under the same delay constraints. What is remarkable is that the gain in performance can be significant. For instance, randomized bucketing can achieve zero min-entropy leakage under the same delay constraints whereas deterministic bucketing cannot. We will illustrate this through a toy example. Suppose there are three secrets $\{0, 1, 2\}$ with the prior $(1/2, 1/3, 1/6)$ and execution times 1, 2, 3. The Pareto-front solutions from the two optimizations are depicted in Fig 7. In particular, the following randomized bucketing strategy δ achieves zero (min-entropy) leakage:

$$\delta(1|0) = \frac{2}{3}, \delta(2|0) = \frac{1}{3}, \delta(1|1) = 1, \delta(2|2) = 1$$

In fact by using Bayes’ rule the posterior is the following

$$p(0|1) = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{3} \cdot 1} = \frac{1}{2} = p(1|1)$$

$$p(0|2) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{6} \cdot 1} = \frac{1}{2} = p(2|2)$$

Therefore, the adversary would have guessed the secret in one try with the same probability of success using the prior or the posterior, i.e., there is no min-entropy leakage.

The delay induced by δ is $\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot 2 \cdot \frac{1}{3} = \frac{2}{3}$. However the delay at which deterministic bucketing achieve zero leakage is $2 \cdot \frac{1}{2} + \frac{1}{3} = \frac{4}{3}$, which is the trivial solution of having a single bucket at the maximum time. Both of these points are visible in Fig 7 (where each Pareto-front touches the x-axis).

$$\begin{aligned} \text{Min: } & \sum_{j=1}^N z_j \\ \text{s. t.: } & \delta(j|i) \geq 0, & i \leq j, j \leq N \\ & \sum_{j=1}^N \delta(j|i) = 1, & i \leq N \\ & z_j \geq p_i \delta(j|i) & i \leq j, j \leq N \\ & \sum_{j=1}^N \sum_{i=1}^j p_i \delta(j|i) (T_j - T_i) \leq \epsilon \end{aligned}$$

Fig. 5. Optimization for randomized bucketing (LP)

$$\begin{aligned} \text{Min: } & \sum_{j=1}^N z_j \\ \text{s. t.: } & \delta(j|i) \in \{0, 1\}, & i \leq j, j \leq N \\ & \sum_{j=1}^N \delta(j|i) = 1, & i \leq N \\ & \delta(j|k) \geq \delta(j|i) & i \leq j, j \leq N \\ & & i \leq k \leq j \\ & z_j \geq p_i \delta(j|i) & i \leq j, j \leq N \\ & \sum_{j=1}^N \sum_{i=1}^j p_i \delta(j|i) (T_j - T_i) \leq \epsilon \end{aligned}$$

Fig. 6. Optimization for Standard bucketing (ILP).

VII. CONCLUSIONS AND FUTURE WORK

We have presented some new advances in designing robust minimal leakage channels. We established the existence of universally optimal channels for a new class of constraints. The question of the extent of existence of universal (measure-invariant) optimality is still an open problem. We expect more

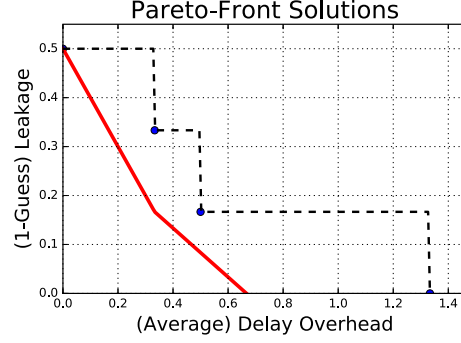


Fig. 7. Comparison between deterministic bucketing (dashed line) and randomized bucketing. The randomized bucketing outperforms the deterministic bucketing with respect to both of the (leakage/utility) objectives. Notably, randomized bucketing can achieve zero min-entropy leakage non-trivially (with average delay of $2/3$) as opposed to the trivial solution of deterministic bucketing of having a single bucket at the end (average delay of $4/3$).

generalized classes of channel design problems that admit a universal (measure-invariant) solution exist. Finding other sufficient and/or necessary conditions for universal (measure-invariant) optimality will be among our future steps. We expect that our foundational results, for instance, the universal convexity of the problem and necessity and sufficiency of KKT conditions would provide the tools to explore this open problem. The bucketing application makes a strong case for randomized channel design. A priority of our further work would be to look into other potential applications and implementations taking into account their nuances.

REFERENCES

- [1] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.
- [2] B. Köpf and M. Durmuth, "A provably secure and efficient countermeasure against timing attacks," in *Proc. of the 22nd Computer Security Foundations Symposium (CSF 2009)*. IEEE, 2009, pp. 324–335.
- [3] B. Köpf and G. Smith, "Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks," in *Proc. of the 23rd Computer Security Foundations Symposium (CSF 2010)*. IEEE, 2010, pp. 44–56.
- [4] X. Cai, R. Nithyanand, T. Wang, R. Johnson, and I. Goldberg, "A systematic approach to developing and evaluating website fingerprinting defenses," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 227–238.
- [5] M. Juarez, M. Imani, M. Perry, C. Diaz, and M. Wright, "Toward an efficient website fingerprinting defense," in *European Symposium on Research in Computer Security*. Springer, 2016, pp. 27–46.
- [6] P. Eckersley, "How unique is your web browser?" in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2010, pp. 1–18.
- [7] 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.
- [8] 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.
- [9] 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 *Proc. of the 13th Workshop on Privacy in the Electronic Society*. ACM, 2014, pp. 73–82.

- [10] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *Proc. of the 27th Computer Security Foundations Symposium (CSF 2014)*. IEEE, 2014, pp. 308–322.
- [11] F. Biondi, A. Legay, P. Malacaria, and A. Wasowski, "Quantifying information leakage of randomized protocols," in *Verification, Model Checking, and Abstract Interpretation*. Springer, 2013, pp. 68–87.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] G. Smith, "On the foundations of quantitative information flow," in *Foundations of software science and computational structures*. Springer, 2009, pp. 288–302.
- [16] M. Khouzani and P. Malacaria, "Relative perfect secrecy: Universally optimal strategies and channel design," in *Proc. of the 29th Computer Security Foundations Symposium (CSF 2016)*. IEEE, 2016, pp. 61–76.
- [17] 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.
- [18] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *Proc. of the 25th Computer Security Foundations Symposium (CSF 2012)*. IEEE, 2012, pp. 265–279.
- [19] F. Biondi, T. Given-Wilson, and A. Legay, "Attainable unconditional security for shared-key cryptosystems," in *Proc. of the 14th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2015)*. IEEE, 2015.
- [20] H. Chen and P. Malacaria, "Studying maximum information leakage using karush-kuhn-tucker conditions," in *Proceedings 7th International Workshop on Security Issues in Concurrency, SECCO 2009, Bologna, Italy, 5th September 2009.*, 2009, pp. 1–15. [Online]. Available: <http://dx.doi.org/10.4204/EPTCS.7.1>
- [21] —, "Quantifying maximal loss of anonymity in protocols," in *Proceedings of the 2009 ACM Symposium on Information, Computer and Communications Security, ASIACCS 2009, Sydney, Australia, March 10-12, 2009*, 2009, pp. 206–217. [Online]. Available: <http://doi.acm.org/10.1145/1533057.1533087>
- [22] G. Doychev and B. Köpf, "Rational protection against timing attacks," in *28th Computer Security Foundations Symposium (CSF 2015)*. IEEE, 2015, pp. 526–536.
- [23] 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)*. Internet Society, 2013.
- [24] S. Fehr and S. Berens, "On the conditional rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6801–6810, 2014.
- [25] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional rényi entropies," in *Information Theoretic Security*. Springer, 2014, pp. 103–121.
- [26] A. Chinchuluun and P. M. Pardalos, "A survey of recent developments in multiobjective optimization," *Annals of Operations Research*, vol. 154, no. 1, 2007.
- [27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [28] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.
- [29] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.
- [30] S. Chen, R. Wang, X. Wang, and K. Zhang, "Side-channel leaks in web applications: A reality today, a challenge tomorrow," in *Security and Privacy (SP), 2010 IEEE Symposium on*. IEEE, 2010, pp. 191–206.
- [31] K. Zhang, Z. Li, R. Wang, X. Wang, and S. Chen, "Sidebuster: automated detection and quantification of side-channel leaks in web application development," in *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 2010, pp. 595–606.