UNIVERSITÀ DEGLI STUDI FIRENZE

DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Searching secrets rationally

Michele Boreale, Fabio Corradi

DISIA WORKING PAPER
2015/05

Searching secrets rationally

Michele Boreale, Fabio Corradi

Universit di Firenze, Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
Viale Morgagni 65, 50134 Firenze (IT).

Abstract

We study quantitative information flow, from the perspective of an analyst who is interested in maximizing its expected gain in the process of discovering a secret, or settling a hypothesis, represented by an unobservable X, after observing some Y related to X. In our framework, inspired by Bayesian decision theory, discovering the secret has an associated reward, while the investigation of the set of possibilities prompted by the observation has a cost. We characterize the optimal strategy for the analyst and the corresponding expected gain (payoff) in a variety of situations. We argue about the importance of *advantage*, defined as the increment in expected gain after the observation if the analyst acts optimally, and representing the value of the information conveyed by Y. We also argue that the proposed strategy is more effective than others, based on probability coverage. Applications to cryptographic systems and to familial DNA searching are examined.

1 Introduction

Broadly speaking, we refer to quantitative information flow (QIF) as the measurement of the quantity of information flowing from a unobservable random variable X to an observable Y. When expressing information as Shannon entropy [12], this quantity is just mutual information, that is, the difference between the prior and conditional entropy of X.

Computer scientists and statisticians have considered QIF from different perspectives. In the context of computer security, QIF measures expected leaks in a probabilistic system, revealing part of the secret X after some Y is observed. For a statistician, QIF corresponds to the expected reduction in uncertainty as the reward for an observation. Attackers, experimental designers and defenders are just few of the very different names assumed by the actors playing in this scene. Here, we take a somewhat neutral perspective, and simply refer to the *analyst* as someone who evaluates how much he can expect from conditioning X on Y, in a scenario involving a cost proportional to the set of possibilities that should be tested, and a reward associated with disclosing the secret.

In the field of quantitative security, Smith [25, 19] has recently considered the problem of providing an adequate QIF measure for a scenario where an analyst is limited to a single guess on the secret. An ATM withdrawing a credit card after two failed attempts at guessing the PIN illustrates the case. In this context, mutual information, which considers the global uncertainty about X before and after observing Y under a $-\log$ scale, was found to be inadequate as a measure of a QIF: in fact, the analyst's guess is now just the mode of X, so his concern is only about $V(X) = \max_x p(x)$ and $V(X|Y) = E_Y(\max_x p(x|y))$, named vulnerability and conditional vulnerability of the system, respectively. Mimicking Shannon entropy, Smith used vulnerability on the $-\log$ scale, thus obtaining an instance of Renyi's entropy called *min-entropy*.

In the present paper, we follow a more general approach to QIF, stemming from the tradition of Bayesian decision theory, as for example expounded in De Groot[16]. The idea is to introduce, for the problem at hand, costs associated with possible actions and a reward for disclosing a secret; then to derive the optimal analyst's action, that is, the one maximizing the overall expected gain. An action is just a set of possibilities that the analyst should test, or somehow further, in order to (hopefully) disclose the secret, given some observable evidence. Min-entropy corresponds to the case where the reward and the costs are fixed in such a way that there is no advantage to go on testing beyond the first, most likely possibility.

In the paper, we first define a general setting from which a gain function and a QIF measure are derived (Section 2). A central role is played by *advantage*, denoted $A(X;Y)$: the difference in expected gain before and after the observation, if the analyst plays an optimal action. This represent the value, for the analyst, of the information that Y conveys about X. We then specialize the analysis by considering a fixed reward α coming from the secret's disclosure and a unit cost for each undertaken attempt (Section 3). In this setting, we derive an optimal strategy to find the secrets to be investigated and characterize the resulting advantage. The strategy is shown to be more effective than both a k-tries strategy with k fixed, and the strategy of investigating secrets up to reaching a fixed probability coverage. Our results are then specialized to the important case of a non informative (uniform) prior on the secret, possibly in the presence of a symmetric or deterministic system (Section 4). In particular, when the reward coming from the secret equals precisely the cost of discovering the secret for sure, we establish that the proposed strategy essentially corresponds to the one derived from the likelihood ratio criterion. We then examine a few applications of the proposed framework, concerning cryptographic systems and the analysis of forensic databases for familial DNA searching (Section 5). Discussion of further and related work concludes the paper (Section 6). Some detailed proofs have been confined to a separate appendix.

2 Set up

We let $\mathcal{X}$ and $\mathcal{Y}$ be finite, nonempty sets of *secrets* and *observables*, respectively. A conditional probability matrix $p_{Y|X} \in [0,1]^{\mathcal{X}\times\mathcal{Y}}$ defines the behaviour of the system under observation, with $p(y|x)$ denoting the probability of observing y when the secret is x. In the terminology of Information Theory, this represents the *channel* through which information flows. A prior probability p_X on $\mathcal{X}$ is assumed; we will drop the index $_X$ whenever X is clear from the context. p_X and the channel matrix $p_{Y|X}$ together give rise to a joint probability distribution on $\mathcal{X}\times\mathcal{Y}$, hence to a pair (X, Y) of input-output random variables, as expected. In many specific contexts, X and Y are not immediately related to one another, but we assume it is possible for the analyst to marginalize out all the unobserved r.v. in the system apart from X. Therefore, both the prior and the conditional probability matrix are assumed to be known to the analyst. We will make freely use of such notational shorthand as $p(y)$ for $\Pr(Y = y)$, $p(x|y)$ for $\Pr(X = x|Y = y)$, and so on, whenever no ambiguity arises as to the involved random variables.

Let $\mathcal{W}$ be a finite, nonempty set of *actions* the analyst can take, possibly after observing Y. Undertaking a certain action under a given state of the world / secret induces a (possibly negative) gain for the analyst, according to a given *gain function* $g : \mathcal{X}\times\mathcal{W} \to \mathbb{R}$. The *expected gain* under p_X and $w \in \mathcal{W}$ and the *maximal expected gain* under p_X are defined respectively as follows:

$$G(X; w) \triangleq E[g(X, w)] = \sum_x g(x, w)p(x) \tag{1}$$

$$G(X) \triangleq \max_{w\in\mathcal{W}} G(X; w)\,. \tag{2}$$

When notationally convenient, we shall use $G(X; w)$ and $G(X)$ interchangeably with $G(p; w)$ and $G(p)$, respectively, thus identifying X by its distribution p_X. In (2), a $w \in \mathcal{W}$ achieving the maximum is called a *Bayes action*. By $w^*(p)$ we indicate a Bayes action, arbitrarily chosen if there is more than one. If no ambiguity arises about p, we abbreviate $w^*(p)$ as w^*.

For $y \in \mathcal{Y}$, let $p(\cdot|y)$ denote the posterior probability distribution on $\mathcal{X}$ given $Y = y$, whenever such an event has nonzero probability, and by $G(X|y) = G(p(\cdot|y))$ the corresponding gain.The *posterior maximal expected gain*, *advantage* (under p_X) and *capacity* of the system are given by:

$$G(X|Y) \triangleq E_y[G(X|y)] = \sum_y p(y)G(X|y) \tag{3}$$

$$A(X; Y) \triangleq G(X|Y) - G(X) \tag{4}$$

$$C \triangleq \sup_{p_X} A(X; Y) \tag{5}$$

where in (3) it is understood that the sum runs over y's of positive probability. Let $\mathcal{P}$ denote the set of all probability distributions over $\mathcal{X}$. General result about expected gain and advantage are the following.

Lemma 1 (convexity) *$G(p)$ is a convex function of p.*

Applying the above lemma and Jensen's inequality, we easily get the following corollary. It says that for the analyst it is always advantageous, on average, to try and guess *after* observing Y rather than before. This is a standard result first published by Raiffa and Schlaifer [23] but also noted by Ramsey in the 1920s.

Corollary 1 *$A(X; Y) \geq 0$ for each $p_X \in \mathcal{P}$. Moreover, if X and Y are independent as random variables – that is, $p_{XY}(x, y) = p_X(x)p_Y(y)$ for each $(x, y) \in \mathcal{X}\times\mathcal{Y}$ – then $A(X; Y) = 0$.*

3 General results on rational analysts

We now instantiate the setup of the previous section to one where an analyst can investigate a set of candidates for the secret, with a *cost* proportional to the size of the set. Moreover, the analyst assigns the secret a certain *value* $\alpha > 0$: this represents the maximal amount of resources the analyst is willing to spend to discover the secret. For notational simplicity, we stipulate that the value of the secret is expressed in cost units – that is, the secret is worth α times the cost of trying a single candidate. This leads to the following definitions.

$$\begin{aligned} \mathcal{W} &\triangleq 2^{\mathcal{X}} && (6) \\ g(x, w) &\triangleq \alpha \cdot 1_{x \in w} - |w| && (7) \end{aligned}$$

where $2^{\mathcal{X}}$ denotes the powerset of $\mathcal{X}$, 1_E is the indicator function of E, which holds 1 if E is true and 0 otherwise, and $| \cdot |$ denotes the cardinality of a set. We begin with characterizing the best strategy of the analyst given a generic prior p_X, that is, the Bayes action corresponding to p_X in the above framework. Let us define the following set

$$w^* \triangleq \{x : p(x) \geq 1/\alpha\}. \qquad (8)$$

For any $w \subseteq \mathcal{X}$, we let $p(w)$ denote $\sum_{x \in w} p(x)$.

Lemma 2 *w^* is a Bayes action. Therefore, $G(p) = G(p; w^*) = \alpha p(w^*) - |w^*|$.*

Proof Consider any $w \subseteq \mathcal{X}$. We have the following.

$$\begin{aligned} G(p, w) &= \sum_{x \in \mathcal{X}} p(x)(\alpha \cdot 1_{x \in w} - |w|) \\ &= \sum_{x \in w} (\alpha p(x) - 1) \\ &= \sum_{x \in w \cap w^*} (\alpha p(x) - 1) + \sum_{x \in w \setminus w^*} (\alpha p(x) - 1) \\ &\leq \sum_{x \in w \cap w^*} (\alpha p(x) - 1) && (9) \\ &\leq \sum_{x \in w \cap w^*} (\alpha p(x) - 1) + \sum_{x \in w^* \setminus w} (\alpha p(x) - 1) && (10) \\ &= \sum_{x \in w^*} (\alpha p(x) - 1) \\ &= G(p, w^*) \end{aligned}$$

where: inequality (9) is justified by the fact that, for $x \notin w^*$, $(\alpha p(x) - 1) < 0$ by definition of w^*; inequality (10) is justified by the fact that, for $x \in w^*$, $(\alpha p(x) - 1) \geq 0$ again by definition of w^*. □

Remark 1 (Bayes action vs. alternative strategies) *In plain words, the above result says that the optimal strategy is obtained by including in w^* candidates from $\mathcal{X}$, considered in descending order of probability, and stopping as soon as the additional cost of the next candidate (1, in α units) equals the expected benefit it brings ($\alpha p(x)$). This is similar to the production stopping rule from Microeconomics: stop producing as soon as marginal revenue equals marginal cost.*

In particular, both the k-try strategy, for k fixed, considered by Smith et al [1] and the α -coverage probability strategy, for α fixed, as in Slooten and Meester [24], are in general sub-optimal, at least when this cost structure is introduced. In both cases, if more than $|w^*|$ secrets are investigated, the net reward for the additional $k - |w^*|$ investigated is negative. On the opposite, if less than $|w^*|$ secrets are considered, the missed $|w^*| - k$ would have achieved a net positive gain.

As a final remark, we note that under the assumption that p_X is known, it is possible to recover the 1-try strategy by fixing any α such that $1/\pi_M < \alpha \leq 1/\pi'_M$, where π_M and π'_M denote the largest and second largest probability values in p_X, respectively.

The above lemma specializes to the following characterization when p_X is uniform. We let $N \stackrel{\triangle}{=} |\mathcal{X}|$.

Corollary 2 *Let p_X be uniform on $\mathcal{X}$. Then the following three cases may arise depending on α.*

- *If $\alpha > N$ then $w^* = \mathcal{X}$ is the only Bayes action, and $G(X) = \alpha - N > 0$.*
- *If $\alpha = N$ then any $w \subseteq \mathcal{X}$ is a Bayes action, and $G(X) = 0$.*
- *If $\alpha < N$ then $w^* = \emptyset$ is the only Bayes action, and $G(X) = 0$.*

For the analyst it is important to estimate the advantage gained by observing Y, over not observing it: this quantifies the value of the information conveyed by Y to him. We study this aspect in the rest of the section. Our first result is a simple remark, saying that the advantage can be decomposed into two parts: one depends on how much the probability mass in the Bayes action gets incremented after the observation; the other on how much the Bayes action shrinks after the observation. A proof is reported in the Appendix. In the sequel, for any y such that $p(y) > 0$, we let w^*_y be the Bayes action associated with the posterior probability distribution $p(\cdot|y)$. Explicitly,

$$w^*_y \stackrel{\triangle}{=} \{x \,:\, p(x|y) \geq 1/\alpha\}\,. \tag{11}$$

Proposition 1 $A(X;Y) = \alpha E_y\left[p(w^*_y|y) - p(w^*)\right] + E_y\left[|w^*| - |w^*_y|\right]$.

Note in particular, that, for a fixed a priori p_X, the maximum of $A(X;Y)$ taken over all possible channels is achieved if the posterior on the secrets degenerates into a point mass function, for every y. Thus, the maximal achievable advantage is $\leq \alpha(1-p(w^*))+(|w^*|-1)$.

After observing Y, an increase in gain can be obtained by observing some other output related to X, say Z, possibly through a different channel $p_{Z|X}$. In other words, we assume that some prior p_{XYZ} is given such that factorizes as $p_{XYZ}(x,y,z) = p(x)p(y|x)p(z|x)$. The whole advantage deriving from observing the pair (Y,Z) can in fact be computed sequentially, as stated by the next proposition. In what follows, we let $G(X|Z,y) \triangleq E_z[G(X|z,y)]$. Clearly, it holds that $G(X|Z,Y) = G(X|Y,Z) = E_y G(X|Z,y)$. We also define $A(X;Z|Y) \triangleq G(X|Y,Z) - G(X|Y)$.

Proposition 2 $A(X;Y,Z) = A(X;Y) + A(X;Z|Y)$.

Proof By definition

$$\begin{array}{rcl} A(X;Y) + A(X;Z|Y) & = & G(X|Y) - G(X) + G(X|Y,Z) - G(X|Y) \\ & = & G(X|Y,Z) - G(X) \\ & = & A(X;Y,Z)\,. \end{array}$$

□

Our last result in this section is a simple formula to estimate advantage. By this, we mean a formula based solely on quantities depending on the system $p_{Y|X}$, plus simple features of the prior p_X. This characterization will be useful when specializing the framework to a number of instances, as we shall see in subsequent sections. We define below a few quantities and sets, also depending on given $y \in \mathcal{Y}$, with the following intuitive meaning. π_M and π_m are the largest and smallest nonzero probability values of the prior. w_y^+, w_y^- represent certain over- and under-approximation of the Bayes action after observing y. They are obtained by weakening (resp. strengthening) the condition $p(y|x) \geq p(y)/(p(x)\alpha)$ by taking into account that

$$S_y \frac{\pi_M}{\pi_m} \;\geq\; \frac{p(y)}{p(x)} \;\geq\; S_y \frac{\pi_m}{\pi_M}\,.$$

Here, S_y is the sum of the entries in the y-column of the conditional probability matrix, while S_y^+, S_y^- are the sums restricted to the rows which enter the over- and under-approximation, respectively, of the Bayes action after y. Formally, we have

$$\begin{array}{rclrclrcl} \pi_M & \triangleq & \max_x p(x) & w_y^+ & \triangleq & \{x : p(y|x) \geq \frac{S_y}{\alpha}\frac{\pi_m}{\pi_M}\} & w_y^- & \triangleq & \{x : p(y|x) \geq \frac{S_y}{\alpha}\frac{\pi_M}{\pi_m}\} \\ \pi_m & \triangleq & \min_{x \in \mathrm{supp}(p)} p(x) & \hat{S}_y^+ & \triangleq & \sum_{x \in w_y^+} p(y|x) & \hat{S}_y^- & \triangleq & \sum_{x \in w_y^-} p(y|x) \\ & & & S_y & \triangleq & \sum_x p(y|x)\,. & & & \end{array}$$

The following result gives upper- and lower-bounds for $G(X|Y)$ based on the above quantities, for an arbitrary prior distribution p_X of full support. The proof is reported in the Appendix.

Proposition 3 *Assume p_X has full support. Then* $\sum_y \left(\alpha\pi_m\hat{S}_y^- - \pi_M S_y |w_y^+|\right) \leq G(X|Y) \leq \sum_y \left(\alpha\pi_M\hat{S}_y^+ - \pi_m S_y |w_y^-|\right)$.

4 Some special cases

4.1 Uniform prior

In case the prior p_X is uniform, we have $\pi_m = \pi_M = 1/N$ in Lemma 3. Note that, in this case, the Bayes action for the adversary after observing $y \in \mathcal{Y}$ is given by

$$w^*_y = w^+_y = w^-_y = \{x : p(y|x) \geq S_y/\alpha\} .$$

As a consequence, the $(\cdot)^+$ and $(\cdot)^-$ sets/quantities defined in the previous section coincide, and we can drop the superscripts from them. The upper and lower bounds given in Proposition 3 coincide too. As a consequence, we have the following characterization of advantage for uniform prior. For convenience we let

$$S^* \triangleq \sum_y \hat{S}_y$$

denote the sum of the entries of the channel matrix that are not less than the threshold S_y/α. The result shows that advantage is proportional to S^*.

Corollary 3 (uniform prior) *Let p_X be uniform.*

- *If $\alpha > N$ then $A(X;Y) = \frac{1}{N}(\alpha S^* - \sum_y S_y|w^*_y|) + N - \alpha$.*
- *If $\alpha \leq N$ then $A(X;Y) = \frac{1}{N}(\alpha S^* - \sum_y S_y|w^*_y|)$.*

Note that the same result could also be obtained from (19) by letting the prior p_X be the uniform distribution.

Remark 2 ($\alpha = N$ and the LR criterion) *The case $\alpha = N$ has a special meaning, since it illustrates a system that is, so to speak, in equilibrium: the cost of discovering the secret with certainty (investigating all the possibilities) equals the revenue coming from discovering the secret. It is interesting to look at the form of the Bayes actions. Again with p_X uniform and $\alpha = N$, we have that*

$$w^*_y \quad = \quad \left\{x : p(y|x) \geq p(y) = \frac{\sum_x p(y|x)}{N}\right\} . \qquad (12)$$

That is, the set w^*_y includes exactly those secrets x such that the likelihood of y under x is at least as big as the average likelihood of y.

Another interesting remark is that, for large N, the inclusion of a secret in the set w^*_y coincides with a decision based solely on the classical likelihood ratio (LR) criterion. To see this, consider any observation y of positive probability and any secret x. According to the LR criterion, secret x receives support by y if $LR(x;y) \geq 1$, that is, if

$$\begin{aligned} LR(x;y) \quad &\triangleq \quad \frac{p(y|x)}{p(y|X \neq x)} = \frac{p(y|x)\frac{N-1}{N}}{\sum_{x' \neq x} \frac{p(y|x')}{\sum_{x'} p(y|x')} p(y)} \\ &= \quad \frac{p(y|x)\frac{N-1}{N}}{\frac{1}{N}\sum_{x' \neq x} p(y|x')} = \frac{p(y|x)}{\frac{1}{N-1}\sum_{x' \neq x} p(y|x')} \geq 1 \ . \end{aligned} \qquad (13)$$

By (12)*, x is in w_y^* if only if*

$$\frac{p(y|x)}{\frac{1}{N}\sum_{x'} p(y|x')} = \frac{p(y|x)}{\frac{1}{N}\sum_{x'\neq x} p(y|x') + \frac{1}{N}p(y|x)} \geq 1. \tag{14}$$

We see that, for N large enough, the two criteria coincide, that is $LR(x;y) \geq 1$ iff $x \in w_y^*$.

4.2 Special channel matrices

An interesting special case is when the conditional probability matrix has columns that are pairwise identical up to a permutation, like for example in the case of a communication protocol (eg. Crowds). Then $|w_y^*|$ does not depend on y, and we let $|w_y^*| = c^*$, for each y. Also note that $\sum_y S_y = N$. We can therefore arrive at a simplification.

Corollary 4 (uniform prior and column-symmetric system) *Let p_X be uniform and assume $p_{Y|X}$ has columns that are pairwise identical up to a permutation.*

- *If $\alpha > N$ then $A(X;Y) = \frac{\alpha}{N}S^* - c^* + N - \alpha$.*
- *If $\alpha \leq N$ then $A(X;Y) = \frac{\alpha}{N}S^* - c^*$.*

Another interesting special case is when the conditional probability matrix defines a deterministic function $f : \mathcal{X} \to \mathcal{Y}$: that is, $p(y|x) = 1$ if and only if $f(x) = y$. Let $c_1, ..., c_K$ be the equivalence classes of $\mathcal{X}$ determined by f, that is, the nonempty inverse images $f^{-1}(y)$ for $y \in \mathcal{Y}$.

Corollary 5 (uniform prior and deterministic system) *Let p_X be uniform and assume $p_{Y|X}$ is deterministic.*

- *If $\alpha > N$ then $A(X;Y) = N - \frac{1}{N}\sum_{j=1}^{K} |c_i|^2$.*
- *If $\alpha \leq N$ then $A(X;Y) = \frac{1}{N}\sum_{|c|\leq\alpha} \alpha|c| - |c|^2$.*

Proof Let us examine the firs part. Under the stated assumptions, it is immediate to check that, for each y of positive probability, letting $c = f^{-1}(y)$, we have: $w_y^* = c$ and $S_y = \hat{S}_y = |w_y| = |c|$. From Corollary 3, the wanted result follows by summing over all y and using some algebra.

The second part is similar: note however that, for a given y, the Bayes action is $w_y^* = c$ if $|c| \leq \alpha$, otherwise it is $w_y^* = \emptyset$. □

5 Applications

5.1 The Crowds anonymity protocol

In the Crowds anonymity protocol, see [22], a set of *honest users* $1, ..., N$ want to exchange messages among one another, but each user wants to hide his identity as a sender from an eavesdropper (attacker). A message initiating from user i is collaboratively and randomly routed through the nodes of a clique network, until it reaches its intended destination. The network's nodes comprise all honest users, plus a number of *corrupted* nodes who collude with the attacker: if any corrupted node receives a message from honest user j, user j is said to have been *detected*. The attacker's task is to identify the user who is the true initiator of the message. Of course, the attacker cannot tell for sure if a detected user j is the true originator of the message or a just a forwarder. This gives rise to a system where $\mathcal{X} = \mathcal{Y} = \{1, ..., N\}$ and $p(j|i)$ is the probability of detecting honest user j, given that honest user i is the true initiator *and* that some user is detected. The resulting matrix has a symmetric form:

$$p_{Y|X} = \begin{bmatrix} \beta & \gamma & \gamma & \cdots & \gamma \\ \gamma & \beta & \gamma & \cdots & \gamma \\ & & \vdots & & \\ \gamma & \gamma & \gamma & \cdots & \beta \end{bmatrix}$$

where the values of β and γ depend on various parameters of the protocol, including: the size of the network, the proportion of corrupted nodes over honest ones, and the *forwarding probability*. The latter is the probability that, upon receiving a message, a honest user forwards it to a randomly chosen node, rather than sending it to its intended recipient. In any case, it holds that $\beta > \gamma$: the probability that the true initiator is detected is (usually, just slightly) higher than that of any other honest user.

Assume now that the prior p on honest users is uniform, and that $\alpha < N$: according to Corollary 3, the best course of action for the adversary, if he cannot observe anything, is just doing nothing, which is realistic in practice. In this case, the advantage of observing the system coincides with the maximal expected gain. We are in the situation of Corollary 4, second item. There are now two possibilities.

- $\beta < 1/\alpha$. In this case, we have $w_j^* = \emptyset$ for each j, so $c^* = 0$, $S^* = 0$, so that $A(X; Y) = 0$. In practice, the value of the secret is too small compared to the effort needed to guess the secret, even after observing the system.

- $\beta \geq 1/\alpha$. We have $w_j^* = \{j\}$ for each j, since it is easy to check that, under the given assumptions, $\gamma < 1/N < 1/\alpha$. As a consequence, $c^* = 1$, $S^* = N\beta$, so that $A(X; Y) = \alpha\beta - 1$. The final benefit for the adversary from guessing the secret after observing the system is a fraction β of the secret's value α. We can make the system less attractive for the attackers by lowering β.

5.2 Cryptosystems

This example is inspired by Shannon's classical information-theoretic treatment of cryptography, as later extended by Hellman [18]. Assume a cryptosystem consists of P possible meaningful plaintext messages, K possible keys and C possible ciphertexts. For any fixed key, enciphering a plaintext results in a unique ciphertext: encryption is deterministic and injective once a key is fixed. For any ciphertext c, consider the set $Z(c)$ of all the plaintext-key pairs (m, k) that give rise to c. By injectivity of encryption, any key appears at most once in $Z(c)$, so $|Z(c)|$ is precisely the number of possible keys that might have been used to generate c, hence a measure of the uncertainty of an attacker about the actual key, given c. Alternatively, $|Z(c)|$ is the cost for an attacker of discovering the secret key, once c is observed. In practice, each candidate key can be tried on another ciphertext c' relative to the same key, to see if it decrypts correctly.

We want to quantify the value for an (ideal) attacker of observing a ciphertext and find a simple lower bound for it. Let then $\mathcal{X}$ be the set of possible meaningful plaintext-key pairs and $\mathcal{Y}$ be the set of possible ciphertexts. Consider the deterministic channel corresponding to the function $f : \mathcal{X} \to \mathcal{Y}$ such that $f((m, k), c) = 1$ if and only if enciphering m with k results in c. Further assume a uniform prior is given on $\mathcal{X}$: this is realistic in certain situations, like for example when the plaintexts are long sequences of letters in a given language, and the key is chosen uniformly at random and independently from the plaintext (see below). Assume, realistically, a reward α such that for each c

$$|Z(c)| \leq \alpha \leq N \triangleq KP\,.$$

Note that the above condition implies that for each ciphertext c, $w^*_c = Z(c)$. In an ideal cryptosystem, $|Z(c)|$ should not vary much depending on c: so let us first assume for simplicity that $|Z(c)| = |Z|$ is a constant not depending on c. Then it is easy to check that

$$|Z| \quad = \quad \frac{PK}{C}$$

consequently, after some algebra we easily obtain

$$A(X; Y) \quad = \quad \alpha - \frac{PK}{C}\,. \tag{15}$$

Let us now drop the assumption that $|Z(c)|$ is constant. Applying the second item of Corollary 5, we obtain

$$A(X; Y) \quad = \quad \alpha - \frac{1}{KP} \sum_c |Z(c)|^2\,. \tag{16}$$

Applying Jensen's inequality to the convex function x^2, we obtain a simple lower bound on the summation in the above expression

$$\sum_c |Z(c)|^2 \quad = \quad C \sum_c |Z(c)|^2/C \;\geq\; C(\sum_c |Z(c)|/C)^2 \quad = \quad \frac{(PK)^2}{C}$$

which when plugged into (16) yields an upper bound for the attacker's advantage similar to the constant case (15)

$$A(X;Y) \quad \leq \quad \alpha - \frac{PK}{C}\,. \tag{17}$$

Assume plaintexts and ciphertexts are blocks of n letters drawn from an alphabet of t symbols, and keys of m bits are used. If the source language has entropy per letter H, there will be (approximately) 2^{nH} equiprobable meaningful plaintexts, $t^n = 2^{n \log t}$ possible ciphertexts and 2^m possible keys, hence (17) becomes

$$A(X;Y) \quad \leq \quad \alpha - 2^{n(H - \log t) + m}\,.$$

To make a concrete case, assuming that the DES cipher, featuring keys of $m = 56$ bits, is employed to encipher sentences in English, with $t = 26$ and $H = 1.5$, we may set: $A(X;Y) \approx \alpha - 2^{-n3.2+56}$. We see that the cost decreases exponentially as the block length n grows, and already with blocks of length around $n = 18$ letters, it is less than 1, meaning that there is nearly no uncertainty as to the key, for the attacker.

In reality, a concrete attacker, with limited computational resources, may not be able to determine $Z(c)$ for each c, so this analysis can be considered as overly pessimistic from a security point of view.

5.3 Familial searching through a forensic DNA database

In several countries a database of familial DNA traits is maintained to give an answer to families looking for a missing relative. The hope is to identify a body occasionally recovered as one of the claimed missing persons. Each family provides the DNA traits of some of their members, and provides the pedigree linking the donors with the missing person. The conditional probability distribution of the claimed relative is the main ingredient to evaluate the probability the body belongs to that family. This process is referred to as familial searching by Evett and Weir [15]. In a different context, the DNA traits of an unknown crime perpetrator, somehow recovered, is compared with the DNA traits of several contibutors included in a data base, in an attempt to establish a relation between the perpetrator and the contributors, or one of their relatives.

The DNA is typed on a number of *loci*, usually located on different chromosomes to exploit independence. At each locus, a genotype, an unordered pair of *alleles*, can be observed. The whole set of alleles pairs of an individual, observed at the considered loci, will be referred to as the *profile* of the individual. The transmission of the alleles along generations is ruled by a transmission mechanism: the first Mendelian law is the simplest possible model. Alleles' probability is almost always estimated by relative frequencies from a sample of the population. Genotypes probabilities for a generic member of the population are derived by population models, via alleles' probability and other parameters tightly related to the specific model. The simplest population model is derived by the Hardy Weinberg conditions and follows a multinomial distribution, see [17]. We need not examine these models in detail, for the time being.

We formally model the problem as follows. The secret random variable X corresponds to n specific identification hypotheses related to the contributors/families, plus the possibility that the perpetrator/corpse is related to the rest of the population: so $\mathcal{X} = \{1, 2, ..., n\} \cup \{Rest\}$. We assume the reference population has size $N > n$, typically with $N \gg n$. Since other identification clues are rarely available, it seems sensible to fix the following distribution, giving the prior probability that a perpetrator/recovered body is related to either any of the donors or to the rest of the population:

$$p(x) \triangleq \begin{cases} \frac{1}{N} & \text{if } x = 1, \dots, n \\ \frac{N-n}{N} & \text{if } x = Rest\,. \end{cases}$$

We let $\mathcal{Y}$ be the set of possible DNA profiles for the perpetrator/recovered body: these might be relative to one locus, or to multiple loci. Finally, we let $p(y|x)$ denote the probability of observing the profile y, given that the perpetrator/body is actually linked to x. We just note that, once a kinship relation between the contributor(s) and the perpetrator/corpse is assumed, $p(y|x)$, for each x and y, is uniquely determined by the chosen transmission and population models (see also Remark 4.)

Application of the set up introduced in the previous sections to the present situation requires a small tweak. Indeed, the element $Rest$ must be filtered out from the set of possible actions, as it makes little sense to investigate the rest of the population as a whole. So the set of actions is now $\mathcal{W} = 2^{\mathcal{X}\setminus\{Rest\}}$. Letting S'_y denote the sum of the elements in the column y of the channel matrix restricted to rows $x \neq Rest$, we have therefore that the Bayes action after observing y can be expressed as

$$w^*_y \;=\; \{x \neq Rest \;:\; p(y|x) \geq \frac{1}{\alpha}(S'_y \;+ (N-n)p(y|Rest))\}. \tag{18}$$

As this expression makes clear, the inclusion of elements in w^*_y is favoured by a high value of the reward α. On the other hand, if $p(y|Rest)$ is high, i.e. if the recovered DNA traits y are fairly common in the population, the number of elements in w^*_y becomes smaller. This effect is enhanced if the proportion of families providing elements for identification, $\frac{n}{N}$, is pretty small, circumventing the illusion to have found interesting clues. Assuming, as it is reasonable, that $\alpha \leq N$, and denoting as we did before by $S^* = \sum_{y,x\in w^*_y} p(y|x)$ the sum of all the matrix entries that are at least as big as S_y/α, advantage takes a simple form

$$A(X;Y) \;=\; \frac{1}{N}(\alpha S^* - (\sum_y S'_y|w^*_y| + (N-n)p(y|Rest)))\,.$$

This expression can be taken to represent the value of the information contained in the database.

Remark 3 *Given the genetic trait y found on a corpse, an interesting question, already posed by [24] for a fixed coverage strategy, is if the probability the body is related to a contributor in w^*_y is greater than the probability it is not. Clearly, if this condition happens* *to be true this encourages further identification activities. We start by noting that*

$$\begin{aligned} p(w_y^*|y) &\propto \sum_{x\in w_y^*} p(y|x)\frac{1}{N} \\ p(Rest|y) &\propto p(y|Rest)\frac{N-n}{N}. \end{aligned}$$

since the probabilities above have the same normalization constant, the condition $p(w_y^*|y) > p(Rest|y)$ happens to be true if

$$\frac{\sum_{x\in w_y^*} p(y|x)}{p(y|Rest)} > N - n.$$

As expected, if n approaches N, the condition becomes easier to be verified. On the other hand, this suggests, once again, to be very cautious when the number of the claimed missing person is a small fraction of the total.

Remark 4 (Computational issues) *Entries in the channel matrix are provided row by row, as the conditional probability of observing each possibly different* DNA *characteristics conditionally to each familial information. Hand calculation is possible but is time consuming and error prone. A commercial algebraical software, DNA-VIEW, see [8], provides single entries of the channel matrix. Alternatively, the entire distribution can be obtained numerically by freely available numerical software dedicated to the problem, such as* Open Familias*, see [14]. In both cases, the choice of the population models and transmission mechanism are limited by the specific software implementations. More freedom can be achieved programming the desired distributions by using a low level language or, as we did, through freely available Bayesian network routines [21]. If the transmission mechanism and the population model realistically take account mutations, population substructure and a level of inbreeding, entries in each row of the matrix assume different values and no equivalence classes arise. A notable exception happens if there is only one familial donor posed on the direct lineage with the missing person and the first Mendelian law is adopted (no mutations). In this case it can be shown, see [11] that, irrespectively of the different number of genotypes in a locus, depending on the number of alleles, only six equivalence classes arise: this typically favours sparse channel matrix and speeds computations.*

6 Conclusion

We have put forward a framework for the analysis of QIF inspired by the Bayesian utility theory. We have argued that the resulting strategies are more cost-effective than strategies, proposed elsewhere in the literature, based on the examination of a fixed fraction of the possibilities or a fixed probability coverage. Applications to a security protocols and to DNA searching have been examined.

Our analysis is confined to the realm of finite spaces, since applications we are mainly interested fall in this category. An extension to secrets and observables defined on reals would deal with the predictive distribution of X before and after the observation of Y, having marginalized out all the nuisance parameters in the model linking them. Bayes actions w^* would resemble HPD intervals, but our proposal, driven by cost and reward functions, does not rely on a fixed coverage probability. For simple models like linear regression, based on the joint multivariate normal, the predictive distribution is available in closed form, in others can be simulated; in any case, it would be interesting to verify if (19) retains its interesting features, that make the advantage dependent on both the increase in probability and the reduced interval length of the predicted secret.

In computer security, there is a growing body of literature on QIF, see e.g. [10, 2, 25, 3, 4, 19, 5, 1, 7] and references therein. In this paper, we have not considered sequential search, in which the analyst can choose his next action based on the results, and updated knowledge, arising from previous observations. This topic has been considered in [6] from a QIF perspective. In this work, however, no cost structure is considered. It would be interesting to cast this aspect too into a Bayesian utility framework.

The *value of information* has been studied in Economics: Marschak noted already in the 1950's that one must clearly distinguish between the amount of information of a source, that can be measured via Shannon entropy, and its value. Cases illustrating this distinction quite sharply can be found in [20]. Marschak's concern was ultimately centered on comparing different information sources (probability distributions, in our terminology), according to the value of information they provide when observed through different channels.

References

[1] M.S. Alvim, K. Chatzikokolakis, C. Palamidessi, G. Smith. Measuring Information Leakage Using Generalized Gain Functions. *IEEE 25th Computer Security Foundations Symposium 2012*, *IEEE Computer Society*: 265-279, 2012.

[2] M. Backes, B. Köpf. Formally Bounding the Side-Channel Leakage in Unknown-Message Attacks. *ESORICS 2008* , *LNCS* 5283: 517-532, Springer, 2008.

[3] C. Braun, K. Chatzikokolakis, C. Palamidessi. Quantitative Notions of Leakage for One-try Attacks. *Proc. of MFPS 2009*, *Electr. Notes Theor. Comput. Sci.* 249: 75-91, 2009.

[4] M. Boreale. Quantifying information leakage in process calculi. *Information and Computation* 207(6): 699-725, 2009.

[5] M. Boreale, F. Pampaloni, M. Paolini. Asymptotic information leakage under one-try attacks. *Proc. of FoSSaCS 2011*, *LNCS* 6604: 396-410, Springer, 2011. Full version to appear on *Mathematical Structures in Computer Science*.

[6] M. Boreale, F. Pampaloni. Quantitative Information Flow under Generic Leakage Functions and Adaptive Adversaries.*Proc. of FORTE 2014*, LNCS 8461:166-181, Springer, 2014.

[7] M. Boreale, M. Paolini. Worst- and Average-Case Privacy Breaches in Randomization Mechanisms. *IFIP TCS 2012*: 72-86, 2012.

[8] C. Brenner. Symbolic kinship program. *Genetics*, 145: 533–542, 1992.

[9] K. Chatzikokolakis, C. Palamidessi, P. Panangaden. Anonimity Protocols as Noisy Channels. *Information and Computation*, 206(2-4): 378-401, 2008.

[10] D. Clark, S. Hunt, P. Malacaria. Quantitative Analysis of the Leakage of Confidential Data. *Electr. Notes Theor. Comput. Sci.* 59(3), 2001.

[11] F. Corradi, F. Ricciardi. Evaluation of kinship identification system based on STR DNA profiles *JRSS C*, 62(5): 649–668, 2013.

[12] T. M. Cover, J. A. Thomas. *Elements of Information Theory*, *2/e*. John Wiley & Sons, 2006.

[13] A. P. Dawid. Coherent measures of discrepancy, uncertainty and dependence, with applications to bayesian predictive experimental design. Department of Statistical Science, University College London. `http://www.ucl.ac.uk/Stats/research/abs94.html`, Tech. Rep. 139, 1998.

[14] T. Egeland, P.F. Mostad. Statistical Genetics and Genetical Statistics. a forensic perspective *Scandinavian Journal of Statistics*, 29(2): 297–307, 2002.

[15] Evett, I. and Weir, B. S. *Interpreting DNA Evidence*. Sinauer Associates, Sunderland, 1998.

[16] M. H. DeGroot. *Optimal Statistical Decisions*, WCL edition. John Wiley & Sons, 2004.

[17] G. H. Hardy. Mandelian proportion in a mixed population. *Science*, 28: 49-50, 1908.

[18] M. E. Hellman. An extension of the Shannon theory approach to cryptography. *IEEE Trans. Info. Theory*, IT-23(3): 289-294, 1977.

[19] B. Köpf, G. Smith. Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. *CSF 2010*: 44-56, 2010.

[20] J. Marschak. *Economics of Information Systems*. Western Management Science Institute. University of California, Los Angeles, November 1969.

[21] K. P. Murphy. The Bayes net toolbox for Matlab *Computing Science and Statistics*, 33: 1024–1034, 2001.

[22] M. Reiter, A. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security* 1(1): 66-92, 1998.

[23] H. Raiffa, R. Schlaifer. *Applied Statistical Decision Theory*. Harward University Press, Boston, 1961.

[24] K. Slooten, R. Meester. Probabilistic strategies for familial DNA searching. *Journal of the Royal Statistical Society series C*, 63(3): 361-384, 2014.

[25] G. Smith. On the Foundations of Quantitative Information Flow. *Proc. of FoSSaCS 2009*, *LNCS* 5504: 288-302, 2009.

A Proofs

Proposition A1 (Proposition 3) *Assume* p_X *has full support. Then* $\sum_y \left(\alpha\pi_m\hat{S}^-_y - \pi_M S_y|w^+_y|\right) \ \leq\ G(X|Y) \ \leq\ \sum_y \left(\alpha\pi_M\hat{S}^+_y - \pi_m S_y|w^-_y|\right)$.

Proof We examine the upper bound only, as the lower bound is symmetric. Fix any y such that $p(y) > 0$. Consider w^*_y, the Bayes action for $p(\cdot|y)$. It is easy to see that $w^-_y \subseteq w^*_y \subseteq w^+_y$. From this, using $p(x|y) = p(y|x)p(x)/p(y)$, it easily follows that

- $p(w^*_y|y) \leq p(w^+_y|y) = \sum_{x\in w^+_y} p(y|x)p(x)/p(y) \leq \hat{S}^+_y\pi_M/p(y)$;
- $|w^*_y| \geq |w^-_y|$.

From the above two inequalities, and by Lemma 2 applied to $p(\cdot|y)$, we have:

$$G(X|y) = G(p(\cdot|y)) = \alpha p(w^*_y|y) - |w^*_y| \leq \alpha\pi_M\hat{S}^+_y/p(y) - |w^-_y|.$$

By averaging the above inequalities on all y of positive probability, and exploiting the following lower bound on $p(y)$

$$p(y) = \sum_{x'} p(y|x')p(x') \geq S_y\pi_m$$

we have

$$G(X|Y) = E_y[G(p(\cdot|y))] = \sum_y p(y)(\alpha p(w^*_y|y) - |w^*_y|) \leq \sum_y \alpha\pi_M\hat{S}^+_y - \pi_m S_y|w^-_y|$$

which the thesis for the considered case. □

Proposition A2 (Proposition 1) $A(X;Y) = \alpha E_y\left[p(w^*_y|y) - p(w^*)\right] + E_y\left[|w^*| - |w^*_y|\right]$.

Proof For any y such that $p(y) > 0$, by plugging (11) into (3), we have

$$G(X|Y) \ = \ \alpha\sum_y p(y)p(w^*_y|y) - \sum_y p(y)|w^*_y|$$

from which, by definition and a suitable rearrangements of the summands

$$\begin{aligned} A(X;Y) &= G(X|Y) - G(X) = \alpha(\sum_y p(y)p(w^*_y|y) - p(w^*)) + (|w^*| - \sum_y p(y)|w^*_y|) \\ &= \sum_y p(y)(\alpha(p(w^*_y|y) - p(w^*)) + |w^*| - |w^*_y|) \\ &= \alpha E_y\left[p(w^*_y|y) - p(w^*)\right] + E_y\left[|w^*| - |w^*_y|\right]. \end{aligned}$$

□