Searching secrets rationally

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Searching secrets rationally

Michele Boreale, Fabio Corradi

Università di Firenze, Dipartimento di Statistica, Informatica, Applicazioni (DiSIA), Viale Morgagni 65, 50134 Firenze, Italy

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 24 November 2015
Accepted 27 November 2015
Available online xxxx

Keywords:
Confidentiality
Quantitative information flow
Decision theory

ABSTRACT

We study quantitative information flow, from the perspective of an analyst who is interested in maximizing its expected gain in the process of learning a secret, or settling a hypothesis, represented by an unobservable X, after observing some Y related to X. In our framework, learning the secret has an associated reward, while the investigation of the set of possibilities prompted by the observation has a cost, proportional to the set's size. Approaches based on probability coverage, or on trying a fixed number of guesses, are sub-optimal in this framework. Inspired by Bayesian decision theory, we characterize the optimal behavior 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 characterize advantage precisely in a number of special but important instances of the framework. 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 [14], 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 can expect a net gain from conditioning X on Y, in a scenario involving a cost proportional to the size of the set of possible guesses, and a reward associated with learning the secret.

In the field of quantitative security, Smith [30] 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 candidates for 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*.

E-mail address: corradi@disia.unifi.it (F. Corradi).

http://dx.doi.org/10.1016/j.ijar.2015.11.010
0888-613X/© 2015 Published by Elsevier Inc.

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 [18]. The idea is to introduce, for the problem at hand, costs associated with possible actions and a reward for learning 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) learn 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 represents 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 learning the secret and a unit cost for each undertaken attempt (Section 3). In this setting, we derive the optimal behavior for the analyst and characterize the resulting advantage. The behavior is shown to be more effective than both a k-tries approach with a fixed k, and the behavior based on trying guesses up to reaching a fixed probability coverage. Our results are then specialized to the important case of a non-informative (uniform) prior on the secrets, 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 learning the secret for sure, we establish that the optimal analyst's behavior essentially corresponds to the one derived from the likelihood ratio criterion. We also consider the maximum advantage that can be obtained over all prior distributions, which is important in security contexts, that is *capacity*. We characterize capacity almost completely in the case of deterministic channels. 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. Setup

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 behavior of the system under observation, with $p(y|x)$ denoting the probability of the observation y when the secret is x. In the terminology of Information Theory, $p_{Y|X}$ represents the *channel* through which information about the secret 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.'s in the system, apart from X. Therefore, both the prior and the conditional probability matrices 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 underlying random variables and distributions.

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:

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

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

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:

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

$$A(X;Y) \overset{\triangle}{=} G(X|Y) - G(X) \tag{4}$$

$$C \overset{\triangle}{=} \sup_{p_X} A(X;Y) \tag{5}$$

where in (3) it is understood that the sum runs over y's of positive probability. General and somewhat standard results about expected gain and advantage are the following. For the sake of completeness, we report their proofs in Appendix A. We let $\mathcal{P}$ be the set of probability distributions on $\mathcal{X}$, seen as a subset of $\mathbb{R}^{|\mathcal{X}|}$.

Lemma 1 *(Convexity). $G(p)$, as a function of p, is convex over $\mathcal{P}$.*

Applying the above lemma and Jensen's inequality, we 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 [27] 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 learn 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.

$$\mathcal{W} \stackrel{\triangle}{=} 2^{\mathcal{X}} \tag{6}$$

$$g(x,w) \stackrel{\triangle}{=} \alpha \cdot 1_{x\in w} - |w| \tag{7}$$

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. Let us emphasize that, as apparent from (7), the analyst's cost is only associated with the action of trying the set of possible guesses, w. Differently from other models considered in the literature (see Section 6), here there is no cost associated with the observation itself. This is sensible in all concrete situations where the observation cost is negligible when compared to the guess-trying cost: think of message interception in a cryptographic system or anonymity protocol like Crowds [28]; or of DNA extraction, whose cost is negligible with respect to checking a correspondence with the crime sample in each investigated family (see Section 5). Remark 1 contains further examples of this situation.

We begin with characterizing the optimal behavior 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

$$w^* \stackrel{\triangle}{=} \{x : p(x) \geq 1/\alpha\}\,. \tag{8}$$

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:

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

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 *(Lack of sequentiality).* In plain words, the above result says that the optimal action is obtained by including in w^* each candidate from $\mathcal{X}$ whose (unitary) cost is not greater than the expected benefit it brings ($\alpha p(x)$). It is important to stress that there is no implication of sequential search for the analyst in this model. This is sensible in those situations where the cost is dominated by $|w^*|$. For a concrete example, consider an offline password attack, where the attacker has got hold of a table of precomputed hashed user passwords, corresponding to w^*: the dominant cost here is the storage requirement for such a table (this is the *capped-guesses* scenario of [22], see related work in Section 6). For another example, consider a scenario where, due to investigative reasons, all suspects prompted by the DNA traits extracted from a crime scene must be tried in a very short time, hence essentially in parallel.

Remark 2 *(Bayes action vs. alternative actions).* The k-tries approach, for fixed k, considered by Smith (see e.g. [30] for the case $k = 1$ and, more generally, [21]) and the coverage model by Slooten and Meester [29], are in general sub-optimal if a cost structure is considered. In both cases, in fact, if more than $|w^*|$ guesses are investigated, the net reward for the additional $k - |w^*|$ investigated guesses, is negative. On the other hand, if less than $|w^*|$ guesses are considered, the missed $|w^*| - k$ ones would achieve a net positive gain.

We note that, under the assumption that p_X is known, it is possible to recover the 1-try approach in our model by fixing any α such that $1/\pi_M < \alpha \le 1/\pi'_M$, 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 $N \overset{\triangle}{=} |\mathcal{X}|$.

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 Appendix A. 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,

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

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

Remark 3. Note that, for a fixed prior p_X, the maximum of $A(X; Y)$ taken over all possible channels is achieved if the posterior on the possible secrets degenerates into a point mass function, for every y. Thus, the maximal achievable advantage is $\le \alpha(1 - p(w^*)) + (|w^*| - 1)$. Likewise, the maximum advantage, hence capacity, for a given channel cannot exceed $\alpha - 1$.

Remark 4. Concerning the decomposition of $A(X; Y)$ given in Proposition 1, we note that the two quantities, $E_y[p(w^*_y|y) - p(w^*)]$ and $E_y[|w^*| - |w^*_y|]$, may well have opposite signs. Generally speaking, for large α, an even modest increase of the covered probability may justify a widening of the search space prompted by an observation. As an extreme case, consider a uniform prior p_X and $\alpha = N$: according to Corollary 2, a prior Bayes action is $w^* = \emptyset$, hence $p(w^*) = |w^*| = 0$. Therefore, for any nontrivial system, we will have $E_y[p(w^*_y|y) - p(w^*)] > 0$ and $E_y[|w^*| - |w^*_y|] < 0$.

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 a joint distribution p_{XYZ} over triples (x, y, z) that factorizes as $p_{XYZ}(x, y, z) = p(x)p(y|x)p(z|x)$. The overall advantage that derives from observing the pair (Y, Z) can in fact be computed sequentially, as stated by the next proposition. In what follows, we let $G(X|Z, y) \overset{\triangle}{=} E_z[G(X|z, y)]$ so it holds that $G(X|Z, Y) = G(X|Y, Z) = E_y G(X|Z, y)$. We also define $A(X; Z|Y) \overset{\triangle}{=} G(X|Y, Z) - G(X|Y)$.

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

Proof. By definition

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

Our last result in this section is a simple formula to estimate $G(X|Y)$. By this, we mean a formula based solely on quantities depending on the system $p_{Y|X}$, plus knowledge of the minimal and maximal probability values in the prior. This characterization is useful because it allows for an estimation of $G(X|Y)$ based on very little knowledge about the prior. From a Bayesian point of view, this corresponds to an intermediate position, in between full knowledge of the prior and non-informativeness. The formula allows for the derivation of results on advantage in a uniform fashion, in a number of special but important instances of the framework that will be considered in the next section.

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, while S_y is the sum of the entries in the y-column of the conditional probability matrix. 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 $p(x|y) \geq 1/\alpha$ as follows. First, focusing on the conditional probability matrix, we note that this condition can be written as $p(y|x) \geq p(y)/(\alpha p(x))$; therefore, taking into account that $p(y) = \sum_{x'} p(y|x')p(x')$, we have

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

Finally, S_y^+, S_y^- are the column sums restricted to those rows which enter the over- and under-approximation, respectively, of the Bayes action after y. Formally, we have

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

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 Appendix A.

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. Special cases

We will examine three instances of the general model that are practically important, and for which we can prove more convenient characterization of advantage.

4.1. Uniform prior

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

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

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 $A(X;Y) = \frac{1}{N}(\alpha S^* - \sum_y S_y|w_y^*|)$.*

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

Remark 5 *($\alpha = N$ 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 learning the secret with certainty (investigating all the possibilities) equals the revenue coming from learning the secret. It is interesting to look at the form of the Bayes actions in this case. Note that with p_X uniform and $\alpha = N$, we have that the condition $p(x|y) \geq 1/\alpha$ is equivalent to $p(y|x) \geq p(y)$. Although the first formulation is more intuitive, we prefer to work here with the second one, which mentions likelihoods, in order to establish a bridge with the LR criterion. Hence we express the posterior Bayes action as

$$w_y^* = \left\{ x : p(y|x) \geq p(y) = \frac{\sum_x p(y|x)}{N} \right\} . \tag{12}$$

That is, the set w_y^* includes exactly those guesses 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 guess 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 guess x. According to the LR criterion, guess x receives support by y if $LR(x; y) \geq 1$, that is, if

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

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

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 (e.g. Crowds). Then $|w_y^*|$ does not depend on y, and we let $|w_y^*| = c^*$, for each y. We also make use of the fact that, in general, $\sum_y S_y = N$. 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 $c_1, \ldots, c_K$, be the equivalence classes of $\mathcal{X}$ determined by f, that is, the nonempty inverse images $f^{-1}(y)$ for $y \in \mathcal{Y}$. Note that $K = |\mathrm{Im}(f)|$.

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 first 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 $w_y^* = \emptyset$. □

4.3. Capacity of deterministic channels

Evaluating capacity is especially important in a security scenario: before deploying a system, a designer may be interested in providing a formal bound to the maximal advantage an attacker could get from the system's observation. This bound should hold irrespective of the probability distribution over the possible secrets, which may be user-specific and not known in advance.

In the deterministic case, we are able to provide an almost complete characterization of capacity, and describe distributions that achieve it. Below, following the notation introduced in the previous subsection, we let $c_1, c_2, \ldots, c_K$ denote the classes determined by the function f, ordered by decreasing cardinality, and corresponding to the observables $y_1, \ldots, y_K$, respectively. To avoid trivialities, we will assume that $\alpha \geq 1$. The proof of the following proposition is based on a somewhat elaborated case analysis, which is reported in Appendix A. We discuss the significance of the results after Corollary 6, which provides a less tight but handier bound for capacity.

Proposition 4. *Assume $p_{Y|X}$ is deterministic and denote by C its capacity.*

1. *If $\alpha \leq K$ then $C = \alpha - 1$. Capacity is achieved by the prior distribution p^*_X such that $p^*_X(x^*_i) = 1/K$, where $x^*_i \in c_i$ is an arbitrary representative of its class, for $i = 1, \ldots, K$.*
2. *If $\alpha \geq N$ then $C = N - \frac{1}{\alpha}\sum_{i=1}^{K}|c_i|^2 - |c_K|(1 - N/\alpha)$. Capacity is achieved by the prior distribution p^*_X such that $p^*_X(x) = 1/\alpha$ if $x \notin c_K$ and $p^*_X(x) = p^*_X(c_K)/|c_K|$ if $x \in c_K$.*

Corollary 6. *Assume $p_{Y|X}$ is deterministic and $\alpha \geq N$. Then $C \leq N(1 - \frac{N}{\alpha K})$.*

Proof. By the previous proposition, part 2, and taking into account that $1 - N/\alpha \geq 0$, we obtain that $C \leq N - \frac{1}{\alpha}\sum_{i=1}^{K}|c_i|^2$. Now write $\sum_{i=1}^{K}|c_i|^2$ as $K\sum_{i=1}^{K}(|c_i|^2/K)$, and apply Jensen's inequality to the (convex) squaring function, to obtain: $\sum_{i=1}^{K}|c_i|^2 \geq K(N/K)^2 = N^2/K$. From this the thesis follows. □

The case $\alpha \leq K$ is the simplest: under the given prior, the value attached to the secret is too small for an analyst to undertake any action before the observation; moreover the observation plainly reveals the secret. The case $\alpha \geq N$ is best understood by looking at the upper bound in Corollary 6 and letting $\alpha = N$. The resulting bound $N(1 - 1/K)$ says that advantage increases as the number of equivalence classes K grows, approaching the ideal value $N - 1$ as K approaches N. In any case, no matter the value of α, the advantage cannot be greater than N: in fact, it approaches N as α approaches infinity. Note that, in general, the distribution achieving capacity is not uniform, but approaches the uniform one as α approaches N from above.

The case $K < \alpha < N$ is not covered by our results. We conjecture that an analysis similar to the case $\alpha \geq N$ could show that $C = \alpha - \sum_{i=1}^{K} n_i^2/\alpha$, with the integers n_i's summing to α (for α integer) and $0 \leq n_i \leq |c_i|$. Such n_i's should be determined by solving an integer quadratic optimization problem. A formal proof of this fact escapes us at the moment.

5. Applications

We will consider three applications of our model: an anonymity protocol, symmetric cryptographic schemes and DNA identification.

5.1. The Crowds anonymity protocol

In the Crowds anonymity protocol [28], a set of *honest users* $1, \ldots, N$ 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 $\mathcal{X} = \mathcal{Y} = \{1, \ldots, N\}$ 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. See [28,11] for details. 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.

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

Assume now that the prior p_X on honest users is uniform, and that $\alpha < N$: according to Corollary 2, 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. Application of this result requires knowledge of the values of c^* and S^*, which we compute below. We distinguish two possible cases:

- $\beta < 1/\alpha$. In this case, we have $w_j^* = \emptyset$ for each j, so $c^* = 0$, $S^* = 0$, so that $A(X;Y) = 0$, by Corollary 4, second item. In practice, the value of learning 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, under the given assumptions, $\gamma < 1/N < 1/\alpha$. As a consequence, $c^* = 1$, $S^* = N\beta$, so that $A(X;Y) = \alpha\beta - 1$, again by Corollary 4, second item. The final benefit for the adversary from learning 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 [20]. Assume a cryptosystem consists of P possible meaningful plaintext messages, L 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 $f(m,k) = c$ if and only if enciphering m with k results in c. Further assume a uniform prior is given on $\mathcal{X}$. This is assumption is realistic in certain situations: for example, when plaintexts are sentences of a given length 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

$$|Z(c)| \leq \alpha \leq N \overset{\triangle}{=} P \cdot L\,.$$

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. We first check that

$$|Z| = \frac{PL}{C}\,.$$

Then after some algebra we can prove obtain

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

Let us now drop the assumption that $|Z(c)|$ is constant. We apply the second item of Corollary 5, and noting that here $N = PL$ and $\sum_{|c| \leq N} |c| = N$ (the size of all classes is by assumption at most N), we obtain

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

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

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

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

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

As an example, assume that blocks of length n, taken from a given source language using an alphabet of t symbols, are enciphered using keys of m bits. If n is large enough and the source language has entropy per letter H, there will be

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

(approximately) 2^{nH} blocks that are legitimate sentences in the language, and they will be (approximately) equiprobable[1]: the set of these blocks will constitute our set of plaintexts. Hence, there are 2^{nH} possible plaintexts, $t^n = 2^{n\log t}$ possible ciphertexts and 2^m possible keys. Therefore (17) becomes

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

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 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 investigators with the DNA traits of some of their members, together with the familial relations (the pedigree) linking the donors with the missing person. For each considered DNA trait, the first task of the investigators consists in deriving the conditional probability distribution of the trait for the missing relative, given the traits of the relatives. This becomes the main ingredient to evaluate the probability that the body under examination is the missing person claimed by that family. This process is referred to as *familial searching* by Evett and Weir [17]. In a different context, the DNA traits of an unknown crime perpetrator, somehow recovered, are compared with the DNA traits of several contributors stored 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 pair of *alleles*, coming from the paternal and maternal lineage, can be observed. The whole set of alleles pairs of an individual, for the considered loci, are the *profile* of the individual. The transmission of the alleles along generations is ruled by known mechanisms: 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 [19]. We need not examine these models in detail here.

We formally model the problem of familial searching 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 $\mathcal{X} = \{1, 2, \ldots, n\} \cup \{Rest\}$. 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:

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

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 pedigree 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 7).

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

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

As this expression should make clear, the inclusion of elements in w^*_y is favored 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 $\alpha \leq N$,

[1] This is Shannon's Asymptotic Equiripartition Property (AEP), also assumed in Hellmann's paper [20].

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

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

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

Remark 6. Given the genetic trait y found on a corpse, an interesting question, already posed by [29] in the case of a fixed coverage approach, is if the probability the body is related to a contributor in w^*_y is greater than the probability it is not. If this condition happens to be true, it encourages further identification activities. We start by noting that

$$p(w^*_y|y) \propto \sum_{x \in w^*_y} p(y|x)\frac{1}{N}$$
$$p(Rest|y) \propto p(y|Rest)\frac{N-n}{N}.$$

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 7 *(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. Paper and pencil calculation is possible but is time consuming and error prone. A commercial algebraical software, DNA-VIEW [10], 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 [16]. 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 [26]. If the transmission mechanism and the population model realistically take into account mutations, population substructure and a level of inbreeding, entries in each row of the matrix take on different values, and no equivalence classes arise. A notable exception occurs if there is only one familial donor in the direct lineage with the missing person, and the first Mendelian law is adopted (no mutations). In this case it can be shown that, irrespectively of the different number of genotypes in a locus, depending on the number of alleles, only six equivalence classes arise [13]: this typically favors sparse channel matrices and speeds up computations.

6. Conclusion and related work

We have put forward a framework for the analysis of QIF inspired by the Bayesian utility theory. Approaches, proposed elsewhere in the literature, based on the examination of a fixed number of guesses or on a fixed coverage of the probability, are suboptimal. Applications to cryptographic schemes and to DNA searching have been considered.

Our analysis is confined to the realm of finite spaces, since applications we are mainly interested in 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 highest posterior density (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 it can be simulated. This topic will be explored in a forthcoming paper. A preliminary investigation has revealed that the interesting property in Proposition 1 is retained: advantage depends on both the variation in probability and interval length of the predicted secret.

In computer security, there is a growing body of literature on QIF. The basic principles have been laid out during the 2000's, see e.g. [12,11,3,30,4,5,21] and references therein. In the last few years, various extensions of the basic model have been considered: to mention but a few, [6] considers the case of one-try (min-entropy) attacks under repeated observations, while the worst- vs. average-case dimension is examined in [9]. Especially relevant to the present paper is the issue of partial secret disclosure [7] and, more generally, of leakage characterization in terms of adversary's (generic) gain, or g-*leakage* [1]. In the last paper, for a generic gain function g, Alvim et al. term the adversary's expected gain as g-*vulnerability*, and define g-leakage by comparing prior and posterior vulnerabilities multiplicatively – that is by taking their log-ratio, rather than their difference as we do here. Our expected gain $G(X)$ might hence be seen as a g-vulnerability. Of course, both

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

our definition and that of [1] are instances of Bayes risk in the sense of Bayesian decision theory, but the concerns are rather different. Indeed, [1] establishes general properties of g-leakage, mainly bounds between min-capacity, g-capacity, and Shannon capacity. On the other hand, the present paper focuses on a specific reward-cost model, for which meaningful results can be given that do not hold for g-leakage in general, nor for min-entropy leakage and its variants in particular.

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 several recent papers from a QIF perspective. In [8] adaptive adversaries are analyzed, however, no cost structure is considered. More general models are in Mardziel et al.'s [23] and [24]. The model in [23] generalizes the classical model in several respects, in particular the defender is allowed to dynamically replace the current secret with a new one, and can adaptively submit inputs to the system, based on feedback from past outputs. [24] enriches this model one step further, by neatly distinguishing the defender loss from the adversary's gain when quantifying leakage. This is relevant when both the defender and the attacker incur some cost. The resulting model is also experimentally validated on a simple example. Although fairly general, the model in [24] is fundamentally different from ours: first, as already discussed, there is no implication of sequentiality in our model; second, and most important, in [24] the attacker incurs a cost upon *observation*, while, as explained in Section 3, in our case a cost are incurred upon performing an *action*, that is, upon testing a set of guesses, either known a priori or prompted by an observation.

In its lack of sequentiality, our cost model is similar to the *capped-guesses* password attack scenario considered by Khouzani et al. in [22]. Assuming that both the defender and the attacker incur some cost when, respectively, picking and guessing a secret, the goal there is to characterize optimal password picking policies for the defender, in the presence of an attacker capable of strategic reasoning. Such policies are characterized via (Nash, maxmin, Stackelberg) equilibria in a game-theoretic setting. The capped-guesses scenario corresponds precisely to the case of an attacker who incurs a cost that depends on the size of his pre-identified list of possible guesses. The attacks are basically brute-force: there is no notion of a posteriori gain, and of consequent advantage/leakage imputable to the functioning of the system. In this respect, the goal of [22] is fundamentally different from ours. Another scenario considered in [22] is *costly-guesses*, where the cost structure implies a sequential search. Attacker's interaction with the system amounts to submitting guesses from a pre-identified list, stopping a soon as a correct guess is made: again, there is no notion of advantage/leakage imputable to observation of the system.

Our results on capacity are related to the recent [2] by Alvim et al. The authors cover several scenarios of capacity for g-vulnerabilities, providing results for additive and multiplicative capacities when universal quantification is allowed over gain functions, over priors, or over both. While the general problem of calculating capacity over all priors for a fixed gain-function remains open, our Proposition 4 can be framed just as the solution of a particular instance of this problem.

The *value of information* has been studied in Economics: Marschak noted already in the 1950's that one must neatly 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 [25]. 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.

Appendix A. Proofs

Lemma A1 *(Lemma 1). $G(p)$, as a function of p, is convex over $\mathcal{P}$.*

Proof. Let $p, q \in \mathcal{P}$, $\lambda \in [0, 1]$ and $r = \lambda p + (1 - \lambda)q$ be a convex combination of p and q. By definition of $G(r)$ and of Bayes action we have the following.

$$\begin{aligned} G(r) &= G(r; w^*(r)) \\ &= \lambda G(p, w^*(r)) + (1-\lambda) G(q, w^*(r)) \\ &\leq \lambda G(p, w^*(p)) + (1-\lambda) G(q, w^*(q)) \\ &= \lambda G(p) + (1-\lambda) G(q)\,. \qquad \square \end{aligned}$$

Corollary A1 *(Corollary 1). $A(X; Y) \geq 0$ for each $X \sim p \in \mathcal{P}$. 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 $A(X; Y) = 0$.*

Proof. First note that p_X can be expressed as a convex combination of the $p(\cdot|y)$, thus: $p_X(\cdot) = \sum_y p_Y(y)p(\cdot|y)$, where as usual y runs over elements of positive probability. Since G is convex by the previous lemma, by Jensen's inequality we get

$$\begin{aligned} G(X) &= G(p_X) \\ &= G\left(\sum_y p_Y(y) p(\cdot|y)\right) \end{aligned}$$

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

$$
\begin{aligned}
&\le \sum_y p_Y(y) G\left(p(\cdot|y)\right) \\
&= G(X|Y)\,.
\end{aligned}
$$

By definition of $A(X;Y)$, this immediately yields the first part of the statement. Assume now that X and Y are independent, so that, for each y of positive probability and for each x, $p(x|y) = p_X(x)$. Then $G(X|Y) = G(X)$, which implies the second part of the statement. □

Proposition A1 *(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

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

Proposition A2 *(Proposition 3). Assume p_X has full support. Then $\sum_y \left(\alpha\pi_m\hat{S}_y^- - \pi_M S_y|w_y^+|\right) \le G(X|Y) \le \sum_y \left(\alpha\pi_M\hat{S}_y^+ - \pi_m S_y|w_y^-|\right)$.*

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 can be seen that $w_y^- \subseteq w_y^* \subseteq w_y^+$. From this, using $p(x|y) = p(y|x)p(x)/p(y)$, it follows that

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

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

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

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

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

we have

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

which is the thesis for the considered case. □

We now come to the proof of Proposition 4. We introduce some additional notation. Once a generic prior p_X has been fixed, we will denote by w^* the induced Bayes action, and by $w_1, \ldots, w_K$ the posterior Bayes actions, determined by the observations of $y_1, \ldots, y_K$, respectively, hence letting p_X implicit. We will also need the following technical lemma.

Lemma A2. *Assume $p_{Y|X}$ is deterministic. Then for any prior distribution, $w^* \cap c_i \subseteq w_i$, for $i = 1, \ldots, K$.*

Proof. Let $x \in w^* \cap c_i$. Then $p(x|y_i) = p(y_i|x)p(x)/p(y_i) = p(x)/p(y_i) = p(x)/p(c_i) \ge 1/\alpha$, where we have exploited $p(y_i|x) = 1$ and $p(x) \ge 1/\alpha$. □

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

Proposition A3 (Proposition 4). *Assume $p_{Y|X}$ is deterministic and denote by C its capacity.*

1. *If $\alpha \le K$ then $C = \alpha - 1$. Capacity is achieved by the prior distribution p^*_X such that $p^*_X(x^*_i) = 1/K$, where $x^*_i \in c_i$ is an arbitrary representative of its class, for $i = 1, \dots, K$.*
2. *If $\alpha \ge N$ then $C = N - \frac{1}{\alpha}\sum_{i=1}^{K}|c_i|^2 - |c_K|(1 - N/\alpha)$. Capacity is achieved by the prior distribution p^*_X such that $p^*_X(x) = 1/\alpha$ if $x \notin c_K$ and $p^*_X(x) = p^*_X(c_K)/|c_K|$ if $x \in c_K$.*

Proof. The first case is obvious because, under the given distribution, $G(X) = 0$ (as $1/K \le 1/\alpha$, so $w^* = \emptyset$ is a Bayes action) and $G(X|y_i) = \alpha - 1$ for each $i = 1, \dots, K$ (as $p(x^*_i|y_i) = 1$ and $w_i = \{x^*_i\}$). Moreover, $\alpha - 1$ is an upper bound to the value of advantage over all possible prior distributions (and, in fact, over all possible channels, see Remark 3).

We consider now case 2, and hence assume $\alpha \ge N$. In what follows, we shall assume without loss of generality that the (prior or posterior) Bayes actions that will be considered include all elements whose probability is $\ge 1/\alpha$. Note that the prior Bayes action w^* cannot be empty, as there must be at least one element x whose probability is $\ge 1/N \ge 1/\alpha$. We first note that, given any prior distribution, taking into account that $p(y_i) = p(c_i)$, it is possible to write advantage in terms of the prior, in the following form

$$A(X;Y) = \sum_{i=1}^{K}(\alpha p(w_i) - p(c_i)|w_i|) - \alpha p(w^*) + |w^*| . \tag{A.1}$$

Fix now any maximizing distribution: later on we will prove that one actually exists. We prove the existence of the desired maximizing distribution by the following successive claims:

1. There is a maximizing distribution for advantage such that $p(c_i) = p(w_i)$ for $i = 1, \dots, K$. To see this, choose, if they exist, $i \in \{1, \dots, K\}$ and $x \in c_i \setminus w_i$ such that $p(x) = \delta$, with $0 < \delta < p(c_i)/\alpha$. Note that w_i cannot be empty, so choose any $x' \in w_i$. We claim that by moving the probability mass δ from x to x', the resulting distribution is not worse off than the original one, in terms of advantage. In fact, taking Lemma A2 into account, one checks that an increment is obtained that can be decomposed, with reference to the terms on the right-hand side of (A.1), as either $\alpha\delta - 0 - 0 + 0$ (if x' does not enter the new w^*) or $\alpha\delta - 0 - \alpha(\delta + p(x')) + 1$ (if x' enters the new w^*; here by assumption $p(x) < 1/\alpha$), which is in both cases nonnegative. By repeating this process many times, we can eventually get a maximizing distribution satisfying the claimed property. Note that, in the resulting distribution, one has $p(\cup_{i=1}^{K} w_i) = 1$. As a consequence, starting from (A.1), it is now possible to rewrite the advantage under this distribution in the following form
$$A(X;Y) = \alpha - \sum_{i=1}^{K} p(w_i)|w_i| - \alpha p(w^*) + |w^*| . \tag{A.2}$$
2. There is a maximizing distribution for advantage such that, within each class c_i, the probability mass is distributed evenly among elements of w_i. Start with a distribution satisfying the property in item 1, and redistribute in a uniform way the probability mass over w_i, within each class. The resulting distribution is not worse off than the original one: this is proven with an analysis based on balancing increments and decrements in the terms of (A.2). One must distinguish the cases when r elements of w_i exit the new w^* or when r elements of w_i enter the new w^*, for some $r \ge 0$. In the first case, the global increment is at least $\alpha r\gamma - r$ for some $\gamma < 1/\alpha$, which is nonnegative. The second case is analyzed similarly.
3. There is a maximizing distribution for advantage satisfying the property in the second item and such that $p(w_i) \le |w_i|/\alpha$ for each $i \ne i_0$, where $|w_{i_0}|$ is minimal among the $|w_i|$'s. Start with a maximizing distribution that satisfies the property in the previous item. Assume there is a class c_i, $i \ne i_0$, whose total probability mass is $p(c_i) = p(w_i) = |w_i|/\alpha + \delta$, with $\delta > 0$. Move the probability mass δ to c_{i_0}, by spreading it uniformly over w_{i_0}. The resulting distribution is not worse off than the original one, as checked by (A.2): one must only be careful in distinguishing the cases $w_{i_0} \subseteq w^*$ and $w_{i_0} \cap w^* = \emptyset$ and, within the latter, whether in the new distribution $w_{i_0} \subseteq w^*$ or not. By repeating this process as many times as needed, we can eventually get a maximizing distribution satisfying the claimed property.
4. There is a maximizing distribution for advantage satisfying the property in the second item and such that $p(w_i) = |w_i|/\alpha$ for each $i \ne i_0$, where $|w_{i_0}|$ is minimal among the $|w_i|$'s. Start with a distribution satisfying the property in the previous item. Assume there is a class c_i such that $p(w_i) = |w_i|/\alpha - \delta$, for some $\delta > 0$. Then one must have $p(w_{i_0}) = |w_{i_0}|/\alpha + \gamma$, with $\gamma \ge \delta$, otherwise the global probability mass would sum to less than 1 (this exploits the fact that $N \le \alpha$). Now move a probability mass δ from c_{i_0} to c_i, by spreading it uniformly over w_i. The proof that the resulting distribution is not worse off than the original one requires a balancing of increments and decrements based on (A.2): in fact, considering that w_i enters the new w^*, the global increment due to the last two terms is $-\alpha(|w_i|/\alpha - \delta) + |w_i| = \alpha\delta \ge 0$. By repeating this many times, we get a distribution with the claimed property. Note that in this distribution $w^* = \cup_{i=1}^{K} w_i$, hence $p(w^*) = 1$. As a consequence, starting from (A.1), it is now possible to rewrite the advantage under this distribution in any of the following equivalent forms

$$
\begin{aligned}
A(X;Y) &= |w^*| - \sum_{i=1}^{K} p(w_i)|w_i| \qquad \text{(A.3)}\\
&= |w^*| - \sum_{i \neq i_0} |w_i|^2/\alpha - (1 - (|w^*| - |w_{i_0}|)/\alpha)|w_{i_0}|\\
&= |w^*| - \sum_{i=1}^{K} |w_i|^2/\alpha - |w_{i_0}|(1 - |w^*|/\alpha)\,. \qquad \text{(A.4)}
\end{aligned}
$$

5. There is a maximizing distribution for advantage satisfying the property in the previous item and such that $w^* = \mathcal{X}$. Assume that there is $x \in c_i \setminus w_i$, that is $x \in c_i$ such that $p(x) = 0$. Again, one can then prove that the probability mass in w_{i_0} must exceed at least $1/\alpha$ the quantity $|w_{i_0}|/\alpha$. By moving the probability mass $1/\alpha$ from w_{i_0} to x, the resulting distribution is not worse off, as it can be seen from equation (A.3). By repeating this many times, we can arrive at a distribution with the claimed property. Note in particular that now $w_i = c_i$ for each $i = 1, \ldots, K$. Under this prior, from (A.4) we can write the resulting advantage as

$$
A(X;Y) = N - \sum_{i=1}^{K} |c_i|^2/\alpha - |c_{i_0}|(1 - N/\alpha)\,. \qquad \text{(A.5)}
$$

6. Now, given that $(1 - N/\alpha) \geq 0$, among all the distributions with the form in (A.5), the one with the smallest c_{i_0}, that is $c_{i_0} = c_K$, is the one that maximizes advantage. This is precisely the distribution described in the statement of the theorem. □

Uncited references

[15]

References

[1] M.S. Alvim, K. Chatzikokolakis, C. Palamidessi, G. Smith, Measuring information leakage using generalized gain functions, in: IEEE 25th Computer Security Foundations Symposium 2012, IEEE Computer Society, 2012, pp. 265–279.

[2] M.S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, G. Smith, Additive and multiplicative notions of leakage, and their capacities, in: IEEE 25th Computer Security Foundations Symposium 2014, IEEE Computer Society, 2014, pp. 308–322.

[3] M. Backes, B. Köpf, Formally bounding the side-channel leakage in unknown-message attacks, in: ESORICS 2008, in: Lect. Notes Comput. Sci., vol. 5283, Springer, 2008, pp. 517–532.

[4] C. Braun, K. Chatzikokolakis, C. Palamidessi, Quantitative notions of leakage for one-try attacks, in: Proc. of MFPS 2009, in: Electron. Notes Theor. Comput. Sci., vol. 249, 2009, pp. 75–91.

[5] M. Boreale, Quantifying information leakage in process calculi, Inf. Comput. 207 (6) (2009) 699–725.

[6] M. Boreale, F. Pampaloni, M. Paolini, Asymptotic information leakage under one-try attacks, in: Proc. of FoSSaCS 2011, in: Lect. Notes Comput. Sci., vol. 6604, Springer, 2011, pp. 396–410. Full version in Math. Struct. Comput. Sci. 25 (2) (2015) 259–291.

[7] M. Boreale, F. Pampaloni, M. Paolini, Quantitative information flow, with a view, in: Proc. of ESORICS 2011, in: Lect. Notes Comput. Sci., vol. 6879, Springer, 2011, pp. 588–606.

[8] M. Boreale, F. Pampaloni, Quantitative information flow under generic leakage functions and adaptive adversaries, in: Proc. of FORTE 2014, in: Lect. Notes Comput. Sci., vol. 8461, Springer, 2014, pp. 166–181. Full version in Log. Methods Comput. Sci. 11 (4–5) (2015) 1–31.

[9] M. Boreale, M. Paolini, Worst- average-case privacy breaches in randomization mechanisms, in: IFIP TCS 2012, 2012, pp. 72–86. Full version in Theor. Comput. Sci. 597 (2015) 40–61.

[10] C. Brenner, Symbolic kinship program, Genetics 145 (1997) 535–542.

[11] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, Anonimity protocols as noisy channels, Inf. Comput. 206 (2–4) (2008) 378–401.

[12] S. Hunt, D. Clark, P. Malacaria, Quantitative analysis of the leakage of confidential data, Electron. Notes Theor. Comput. Sci. 59 (3) (2001) 1–14.

[13] F. Corradi, F. Ricciardi, Evaluation of kinship identification systems based on STR DNA profiles, J. R. Stat. Soc., Ser. C, Appl. Stat. 62 (5) (2013) 649–668.

[14] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd ed., John Wiley & Sons, 2006.

[15] A.P. Dawid, Proper measures of discrepancy uncertainty and dependence with applications to predictive experimental design, Tech. rep. 139, Department of Statistical Science, University College London, 1994, http://www.ucl.ac.uk/Stats/research/abs94.html.

[16] T. Egeland, P.F. Mostad, Statistical genetics and genetical statistics: a forensic perspective, Scand. J. Stat. 29 (2) (2002) 297–307.

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

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

[19] G.H. Hardy, Mandelian proportion in a mixed population, Science XXVIII (1908) 49–50.

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

[21] B. Köpf, G. Smith, Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks, Courr. SF 2010 (2010) 44–56.

[22] M.H.R. Khouzani, P. Mardziel, C. Cid, M. Srivatsa, Picking vs. guessing secrets: a game-theoretic analysis, Technical report 2015, CoRR abs/1505.02325. Short version in CSF 2015.

[23] P. Mardziel, M.S. Alvim, M.W. Hicks, M.R. Clarkson, Quantifying information flow for dynamic secrets, in: IEEE Symposium on Security and Privacy, 2014, pp. 540–555.

[24] P. Mardziel, M.S. Alvim, M.W. Hicks, Adversary gain vs. defender loss in quantified information flow, in: Unofficial Proceedings of Foundations of Computer Security Workshop, 2014.

[25] J. Marschak, Economics of information systems, Western Management Science Institute, University of California, Los Angeles, November 1969.

[26] K.P. Murphy, The Bayes net toolbox for Matlab computing science and statistics, in: Proceedings of the Interface, in: Comput. Sci. Statist., vol. 33, 2001, pp. 1024–1034.

Please cite this article in press as: M. Boreale, F. Corradi, Searching secrets rationally, International Journal of Approximate Reasoning (2015), http://dx.doi.org/10.1016/j.ijar.2015.11.010

[27] H. Raiffa, R. Schlaifer, Applied Statistical Decision Theory, Harward University Press, Boston, 1961.
[28] M. Reiter, A. Rubin, Crowds: anonymity for web transactions, ACM Trans. Inf. Syst. Secur. 1 (1) (1998) 66–92.
[29] K. Slooten, R. Meester, Probabilistic strategies for familial DNA searching, J. R. Stat. Soc., Ser. C, Appl. Stat. 63 (3) (2014) 361–384.
[30] G. Smith, On the foundations of quantitative information flow, in: Proc. of FoSSaCS 2009, in: Lect. Notes Comput. Sci., vol. 5504, Springer-Verlag, Berlin, 2009, pp. 288–302.