Journal of Computer and System Sciences ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Quantifying leakage in the presence of unreliable sources of information

Sardaouna Hamadou [a,*], Catuscia Palamidessi [a,1], Vladimiro Sassone [b,2]

[a] INRIA and LIX, Ecole Polytechnique, France
[b] University of Southampton, United Kingdom

ARTICLE INFO

Article history:
Received 19 August 2016
Accepted 19 August 2016
Available online xxxx

Keywords:
Information hiding
Quantitative information flow
Belief combination
Probabilistic models
Uncertainty
Accuracy

ABSTRACT

Belief and min-entropy leakage are two well-known approaches to quantify information flow in security systems. Both concepts stand as alternatives to the traditional approaches founded on Shannon entropy and mutual information, which were shown to provide inadequate security guarantees. In this paper we unify the two concepts in one model so as to cope with the frequent (potentially inaccurate, misleading or outdated) attackers' side information about individuals on social networks, online forums, blogs and other forms of online communication and information sharing. To this end we propose a new metric based on min-entropy that takes into account the adversary's beliefs.

1. Introduction

Protecting *sensitive* and *confidential* data is becoming increasingly important in many fields of human activities, such as electronic communication, auction, payment and voting. Many protocols for protecting confidential information have been proposed in the literature. In recent years the frameworks for reasoning, designing, and verifying these protocols have considered probabilistic aspects and techniques for two reasons. First, the data to be protected often range in domains naturally subject to statistical considerations. Second and more important, the protocols often use randomised primitives to obfuscate the link between the information to be protected and the observable outcomes. This is the case, e.g., of the DCNets [1], Crowds [2], Onion Routing [3], and Freenet [4].

From the formal point of view, the *degree of protection* is the converse of the *leakage*, i.e. the amount of information about the secrets that can be deduced from the observables. Early approaches to information hiding in literature were the so-called *possibilistic approaches*, in which the probabilistic aspects were abstracted away and replaced by non-determinism. Some examples of these approaches are those based on *epistemic logic* [5,6], on *function views* [7], and on *process calculi* [8,9]. Subsequently, however, it has been recognized that the possibilistic view is too coarse, in that it tends to consider as equivalent randomized obfuscation methods that have very different degrees of protection.

The *probabilistic approaches* became therefore increasingly more popular. At the beginning they were investigated mainly at their strongest form of protection, namely to express the property that the observables reveal no (quantita-

* Corresponding author.
E-mail address: sardaouna.hamadou@gmail.com (S. Hamadou).

[1] The work of Catuscia Palamidessi was partially supported by the INRIA Large Scale Initiative CAPPRIS (Collaborative Action for the Protection of Privacy Rights in the Information Society).
[2] The work of Vladimiro Sassone was partially supported by the project Horizon 2020: SUNFISH (Grant No. 644666).

http://dx.doi.org/10.1016/j.jcss.2017.03.013
0022-0000/

tive) information about the secrets (*strong anonymity*, *no interference*) [1,6,10]. Such strong property, however, is almost never achievable in practice. Hence, weaker notions of protection started to be considered. We mention in particular Rubin and Reiter's concepts of *possible innocence* and of *probable innocence* [2] and their variants explored in [11]. These are, however, still true-or-false properties. The need to express in a quantitative way the degree of protection has then lead naturally to explore suitable notions within the well-established fields of *Information Theory* and of *Statistics*.

Concepts from Information Theory [12] have proved quite useful in this domain. In particular, the notion of noisy channel has been used to model protocols for information-hiding, and the flow of information in programs. The idea is that the input $s \in \mathcal{S}$ of the channel represents the information to be kept secret, and the output $o \in \mathcal{O}$ represents the observable. The noise of the channel is generated by the efforts of the protocol to hide the link between the secrets and the observable, usually by means of randomised mechanisms. Consequently, an input s may generate several different outputs o, according to a conditional probability distribution $p(o \mid s)$. These probabilities constitute the *channel matrix* $\mathcal{C}$. Similarly, for each output there may be several different corresponding inputs, according to the converse conditional probability $p(s \mid o)$ which is linked to the above by the Bayes law: $p(s \mid o) = p(o \mid s)\, p(s)/p(o)$. The probability $p(s)$ is the *a priori* probability of s, while $p(s \mid o)$ is the *a posteriori* probability of s, after we know that the output is o. These probability distributions determine the *entropy* and the *conditional entropy* of the input, respectively. They represent the uncertainty about the input, before and after observing the output. The difference between entropy and conditional entropy is called the *mutual information* and expresses how much information is carried by the channel, i.e. how much uncertainty about the input we lose by observing the output (i.e., equivalently, how much information about the input we gain by observing the output).

Even though several notions of entropy have been proposed in Information Theory, Shannon's is by far the most famous of them, due to its relation with the *channel's rate*, i.e., the speed by which information can be transmitted accurately on a channel. Consequently, there have been various attempts to define the degree of protection by using concepts based on Shannon entropy, notably mutual information [13–16] and the related notion of capacity, which is the supremum of the mutual information over all possible input distributions, and which therefore represents the worst case from the point of view of security [17–19].

A refinement of the above approaches came from the ideas of integrating the notions of extra knowledge and belief [20–22]. The idea is that the gain obtained by looking at the output should be relative to the possible initial knowledge or belief that an attacker may have about the secret. For instance, assume that in a parliament composed by m Labourists and n Conservatives, m members voted against a proposal to eliminate the minimum wage. Without any additional knowledge it is reasonable to believe that all Labourists voted against. If however we came to know that exactly one Conservative voted against, then it is more reasonable to believe that the most liberally-inclined Conservative voted against, and the least liberally-inclined Labourist voted in favour. In this case, the a posteriori belief is likely to be much more accurate than the a priori one, and the gain obtained using the knowledge about MPs' relative positioning on the left-to-right scale is much larger than the one computed as difference of entropies. Consequently, [22] proposed to define the protection of a system in terms of the difference (expressed in terms of Kullback–Leibler divergence) between the accuracy of the a posteriori belief and the accuracy of the a priori one.

Another criticism to the Shannon-entropy-based approach came from Smith, who argued that it is not very suitable to model information leakage in the typical scenario of protocol attacks, where the adversary has only a limited number of tries to guess the value of the secret [23]. In such a scenario, the natural measure of the threat is the *probability* that the adversary guesses the right value. The case of "one-try only" was dubbed by Smith *vulnerability of the secret*. Shannon entropy, on the other hand, represents the expected number of attempts that an adversary has to make to discover the secret, assuming that there is no limit to such number, and that the adversary can narrow down the value by probing properties of the secret. Smith gave an example of two programs whose Shannon's mutual information is about the same, yet the probability of making the right guess after having observed the output is much higher in one program than in the other. In a subsequent paper [24], Smith proposed to define the leakage in terms of a notion of mutual information based on Rényi *min-entropy* (the logarithm of the vulnerability), which captures the case of an adversary disposing of one single try. Subsequent approaches going under the name of g-leakage have extended the analysis to multiple tries, and to the case in which each guess is associated with a gain (or loss) which depends on the level of approximation [25–27]. The min-entropy approach remains however the canonical framework, not only for its simplicity, but also because the worst-case min-leakage (aka min-capacity) has been proved to be an upper bound to the g-leakage [25].

In [28] the authors extended the vulnerability model of [24] in the context of the Crowds protocol for anonymous message posting to encompass the frequent situation where attackers have extra knowledge. They pointed out that in Crowds the adversary indeed has extra information (viz., the target servers) and assumed that she knows the correlation between that and the secret (viz., the users' preferences for servers). They proved that in such scenarios anonymity is more difficult to achieve.

In our opinion, a fundamental issue remains wide open: the need to measure and account for the *accuracy* of the adversary's extra knowledge. Indeed, [28] assumes that the adversary's extra information is accurate, and such an assumption is generally not warranted. Inaccuracy can indeed arise, e.g. from people giving deliberately wrong information, or simply from outdated data. As already noticed in [22] there is no reason in general to assume that the probability distributions the attacker uses are correct, and therefore they must be treated as *beliefs*.

Please cite this article in press as: S. Hamadou et al., Quantifying leakage in the presence of unreliable sources of information, J. Comput. Syst. Sci. (2017), http://dx.doi.org/10.1016/j.jcss.2017.03.013