Does Additional Information Always Reduce Anonymity?

Claudia Diaz Carmela Troncoso George Danezis
K.U.Leuven, ESAT/COSIC,
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium.
firstname.secondname@esat.kuleuven.be

ABSTRACT

We discuss information-theoretic anonymity metrics, that use entropy over the distribution of all possible recipients to quantify anonymity. We identify a common misconception: the entropy of the distribution describing the potential receivers does not *always* decrease given more information. We show the relation of these a-posteriori distributions with the Shannon conditional entropy, which is an average over all possible observations.

Categories and Subject Descriptors: C.4 [Computer System Organizations]: Performance of Systems: Measurement techniques

General Terms: Algorithms, Measurement, Theory

Keywords: Anonymity Metrics, Entropy, Mix, User Profiles

1. INTRODUCTION

The most widely accepted definition of anonymity was given by Pfitzmann and Hansen in [13]: "anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*." The *anonymity set* is "the set of all possible subjects who might cause an action." In other words, subjects are more anonymous as they can hide in a larger crowd.

The adversary of an anonymity system can typically obtain a probability distribution linking an action to all possible subjects who may be related to it. The adversary's uncertainty on the identity of the subject behind an action depends on the number of subjects in the anonymity set, but also on how the probability distribution looks like: as subjects appear more equally likely to be related to the action, the adversary has less information on who might be the real subject linked to it.

The information-theoretic concept of *Shannon entropy* [16] (or simply "entropy") is a measure of the uncertainty associated with a random variable. Technical measures of anonymity [10, 15] are based on the entropy of the probability distribution linking an action to all possible subjects who may be related to it, and they give a measure of the uncertainty of the attacker. Shannon entropy has been very useful in the evaluation [8, 9] of mix-based [2] communication. Metrics based on this entropy have also been proposed to measure the anonymity of profiled users [3, 7].

However, some aspects of entropy-based anonymity metrics are not yet well understood, such as the combination of several sources of information. It has been claimed that an adversary with access to more information is *always* able to reduce anonymity [4]. In this paper, we show that the combination of user profile information with observations at the communication layer does not *necessarily* lead to a reduction of the attacker's uncertainty.

The key misunderstanding stems from the confusion of the attacker's uncertainty in a given scenario with Shannon's *conditional entropy* [16]. We explain here that the attacker's uncertainty is given by the entropy of a conditional probability distribution: the probability that a message was sent to each possible recipient, given a user sending profile and a concrete observation of the communication layer. The entropy of this probability distribution is ***not*** the *conditional entropy*. Therefore, known properties of the *conditional entropy* do not apply to the attacker's uncertainty.

We present in the next section the definition of Shannon's entropy, and introduce information-theoretic anonymity metrics. In Sect. 3 we describe the applications of these metrics to mixes and user profiles. Section 4 explains why the combination of two sources of information does not necessarily decrease the uncertainty, as previously claimed. And finally, we present our conclusions in Sect. 5.

2. INFORMATION-THEORETIC ANONYMITY METRICS

The key concept behind information-theoretic anonymity metrics [10, 15] is *Shannon entropy* [16]. Entropy gives a measure on the uncertainty of a random variable. It increases with the number of non-zero possible outcomes of the random variable, and with the uniformity of the distribution. It is defined as follows: "Let X be a discrete random variable taking a finite number of possible values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively, such that $p_i > 0$ for $i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \quad ." \tag{1}$$

Information-theoretic anonymity metrics use the entropy of

WPES'07, October 29, 2007, Alexandria, Virginia, USA.

the probability distribution that links actions to subjects as a measure of their anonymity. They have been widely used to evaluate anonymous systems, and Steinbrecher and Köpsell [17] noted that they can be used more generally to model unlinkability.

Proposals using other flavors of entropy have followed. Tóth et al. [18] argue that Shannon entropy may not provide relevant information to some users, as it considers the average and not the worst case scenario for a particular user. They suggest using a local anonymity measure computed from min-entropy and max-entropy. More recently, Clauß and Schiffner [4] have proposed Rényi entropy [14] as a generalization of Shannon, min- and max-entropy-based anonymity metrics. We note that Shannon entropy expresses the uncertainty in bits of the outcome of a discrete random variable; however, the interpretation of the other sorts of entropy is unclear.

3. APPLICATIONS OF INFORMATION-THEORETIC ANONYMITY METRICS

3.1 Mixes

The *mix*, first proposed by Chaum [2], is a router that hides the correspondence between inputs and outputs. In order to do so, the mix applies a cryptographic transformation (encryption or decryption) to input messages, in order to change their appearance. The mix also delays and reorders messages, so that outputs cannot be trivially correlated with inputs based on timing information.

The information-theoretic anonymity metrics in [10, 15] have been most useful in the evaluation of mix-based anonymous communication systems. An adversary who knows the internal reordering algorithm of the mix and can observe the messages going in and out of it, is able to compute the probability distribution linking each input to all its possible outputs and vice versa [8, 9].

Let us consider a threshold pool mix with threshold T and pool P. This mix collects up to T messages, places them in the pool (internal memory of the mix) after applying a cryptographic transformation, and forwards $T - P$ of them (in random order). The mix keeps P messages in the pool for the next round.

Let i be the number of rounds a message $\mathcal{M}$ spends in the mix before being sent out to its recipient ($i = 1$ when the message is sent out in the same round it arrived). The adversary observes $\mathcal{M}$ being input to the mix, and the recipients of the subsequent messages output by the mix. Let the discrete random variable X with probability mass function $P_X(i) = \Pr(X = i)$ express the probability of $\mathcal{M}$ spending i rounds in the mix, and thus being received by the i-th recipient. The entropy of X measures the uncertainty of the adversary on the receiver of $\mathcal{M}$. For this mix, the *effective anonymity set size* [15] given by the entropy $H(X)$, computed as in (1), is:

$$H(X) = -\sum_{i=1}^{\infty} P_X(i) \log_2 P_X(i) \ . \qquad (2)$$

3.2 User profiles

Information-theoretic anonymity metrics have been mostly used to evaluate mix-based anonymous communication systems, although similar models are used to quantify anonymity in application-layer contexts, such as user profiles [3, 7]. These profiles consist of a set of attributes that characterize the user (e.g., the user speaks English) or his behavior (e.g., the user communicates frequently with a fixed set of other subjects). One example of attacks that lead to a profile of users' communication patterns are *disclosure attacks*.

Disclosure attacks [1, 5, 6, 11, 12] are one of the most powerful family of attacks on mix-based anonymous communication. The assumption behind these attacks is that users repeatedly send messages to a subset of all possible recipients. A user Alice is modeled as sending messages to a set of recipients with a probability distribution defined by the discrete random variable Y, such that Alice sends a message to recipient y_i with probability $P_Y(y_i) = \Pr(Y = y_i)$, and $\sum_i P_Y(y_i) = 1$.

The goal of disclosure attacks is to uncover the probabilities $P_Y(y_i)$ through the observation of many rounds of communication of Alice. Once the attacker has succeeded, when he sees Alice sending a message $\mathcal{M}$, his uncertainty on the recipient of $\mathcal{M}$ is given by:

$$H(Y) = -\sum_{i} P_Y(y_i) \log_2 P_Y(y_i) \ .$$

4. COMBINING INFORMATION FROM THE COMMUNICATION AND APPLICATION LAYERS

A number of research papers [4, 12, 17] propose combining information obtained from the application and communication layers in order to reduce anonymity. Clauß and Schiffner [4] argue that the entropy of the distribution expressing the uncertainty of the attacker on Alice's choice, given Alice's profile *and* a particular trace at the communication layer, *must* necessarily be lower than the uncertainty of an attacker who has access only to the profile information or to the communication layer observation (quote from [4]):

> "Let X and Y be probability distributions of the application layer and the network layer. One can measure anonymity $H(X)$ and $H(Y)$. As specified roughly in 3.3, the attacker could build a combined model by introducing the circumstances of communication as attributes in the application layer model. Due to the fact that new information can only reduce the cardinality of the set of suspects the resulting probability distribution gets more unequal, i.e., entropy decreases."

We now show with a counterexample that this is not necessarily the case; i.e., the attacker's uncertainty can go up in some scenarios. Individual combinations of profile information and network layer observations may lead to higher entropies without contradicting Shannon's result, as the attacker's uncertainty being measured is not given by Shannon's *conditional entropy*.

4.1 Couterexample

Let us consider that user Alice has a known profile of communicating with five other users (B, C, D, E and F) with a probability distribution $P_Y(y_i)$ such that:

$$P_Y(\mathrm{B}) = 0.04, P_Y(\mathrm{C}) = 0.06, P_Y(\mathrm{D}) = 0.1,$$

$$P_Y(\mathrm{E}) = 0.3, P_Y(\mathrm{F}) = 0.5$$

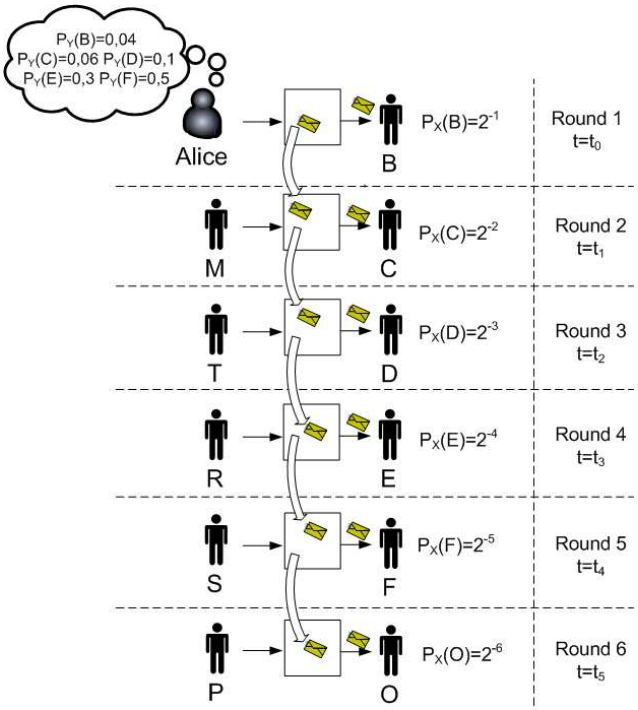

Figure 1: Example scenario.

Alice sends her messages through a threshold pool mix with threshold $T = 2$ and pool $P = 1$. At time t_0, user Alice sends a message $\mathcal{M}$ through the mix to a receiver y_i chosen according to the distribution $P_Y(y_i)$. Based on Alice's profile information, the anonymity $\mathcal{M}$'s recipient is:

$$H(Y) = -\sum_{i=1}^{5} P_Y(y_i) \log_2 P_Y(y_i) = 1.78 \ . \quad (3)$$

In order to further identify the recipient of $\mathcal{M}$, the adversary monitors the outputs of the mix from t_0 on, and sees that the first five outputs go to B, C, D, E and F (in this order), as shown in Fig. 1. For simplicity, we assume that the following recipients (O, etc.) are people to whom Alice never sends messages. Taking into consideration the mix only, the adversary derives the probability $P_X(i) = 2^{-i}$ of Alice's message spending i rounds in the mix before being sent to its recipient [8, 9]. From the observation of the mix outputs (without taking into account Alice's sending profile), the uncertainty of the attacker on Alice's recipient is:

$$H(X) = -\sum_{i=1}^{\infty} P_X(i) \log_2 P_X(i) = 2.$$

Note that the mix does not take into account the identity of Alice's recipient when choosing which message to send next. This means that the probabilities $P_Y(y_i)$ and $P_X(i)$ describing Alice's choice and the mix' choices respectively, are *independent*. Given that the attacker observes the particular sequence of outputs shown in Fig. 1, and taking into account the known Alice profile information, there are five possibilities on who received Alice's message:

1. Alice's message was for B ($P_Y(\mathrm{B}) = 0.04$), and it was immediately sent to B by the mix ($P_X(1) = 2^{-1}$).
2. Alice's message was for C ($P_Y(\mathrm{C}) = 0.06$), it was kept by the mix for one round and then delivered to C ($P_X(2) = 2^{-2}$).
3. Alice's message was for D ($P_Y(\mathrm{D}) = 0.1$), kept for two rounds and delivered ($P_X(3) = 2^{-3}$).
4. Alice's message was for E ($P_Y(\mathrm{E}) = 0.3$), kept for three rounds and delivered ($P_X(4) = 2^{-4}$).
5. Alice's message was for F ($P_Y(\mathrm{F}) = 0.5$), kept for four rounds and delivered ($P_X(5) = 2^{-5}$).

We define a random variable Z that combines both Alice's profile and the communication layer observation. Its entropy $H(Z)$ represents the attacker's uncertainty on the recipient of Alice's message. In this case, Z takes the value $\{z_i\} = \{\mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}\}$ with probability $P_Z(z_i)$:

$$P_Z(z_i) = \frac{P_Y(y_i)P_X(i)}{\sum_j P_Y(y_j)P_X(j)} \ . \quad (4)$$

We obtain the following probability mass function for Z, and its entropy:

$$P_Z(\mathrm{B}) = 0.25, P_Z(\mathrm{C}) = 0.18, P_Z(\mathrm{D}) = 0.15,$$

$$P_Z(\mathrm{E}) = 0.23, P_Z(\mathrm{F}) = 0.19,$$

$$H(Z) = 2.3 \ . \quad (5)$$

According to [4], combining profile information and communication observations must diminish anonymity, as taking more information into account can only reduce the uncertainty of the adversary. This means that the entropy of the combined distribution Z must necessarily be lower than the entropies of the individual distributions X and Y.

However, in this case the entropy $H(Z)$ of the combined distribution, which takes into account both the profile information and the network layer observation, is higher than the entropy $H(Y)$ of the profile information, and that the entropy $H(X)$ of the network observation ($H(X) = 2$, $H(Y) = 1.78$, and $H(Z) = 2.3$). This means that an attacker with access to more information is more uncertain on the recipient of the message than an attacker who only knows Alice's profile but cannot observe the outputs of the mix; and than an attacker who observes the outputs of the mix but does not know Alice's sending profile. This would indeed be a surprising result if $H(Z)$ represented the conditional entropy $H(Y|X)$. But it does not, as we explain in the next section.

4.2 Relationship to conditional entropy

Let us call X the discrete random variable that describes all possible mix output observations $\{x_j\}$, which happen with probability $\Pr(x_j)$. Let Y describe Alice's sending profile: she sends to recipient y_i with probability $\Pr(y_i)$, such that $\sum_i \Pr(y_i) = 1$.

Shannon defines the *conditional entropy* [16] of Y given X, $H(Y|X)$, as "the *average* of the entropy of Y for each value of X, weighted according to the probability of getting that particular x_j. That is:"

$$H(Y|X) = -\sum_{i,j} \Pr(y_i, x_j) \log_2 \Pr(y_i|x_j) \ .$$

Taking into account that:

$$\Pr(y_i, x_j) = \Pr(x_j) \Pr(y_i|x_j) \ ,$$

the conditional entropy $H(Y|X)$ is:

$$H(Y|X) = -\sum_j \Pr(x_j) \sum_i \Pr(y_i|x_j) \log_2 \Pr(y_i|x_j) \ ,$$

Let Z represent the conditional probability of Alice's choice being $z_i = y_i$, given a particular set of mix output observations x_j, such that:

$$\Pr(z_i) = \Pr(y_i|x_j) \ .$$

Given an observation x_j, the uncertainty of the attacker is given by the entropy $H_j(Z)$:

$$H_j(Z) = -\sum_i \Pr(y_i|x_j) \log_2 \Pr(y_i|x_j) \ .$$

Therefore, the conditional entropy $H(Y|X)$ can be expressed as:

$$H(Y|X) = -\sum_j \Pr(x_j) H_j(Z) \ .$$

As we can see, $H(Y|X)$, the weighted average of all possible entropies obtained by the adversary, is **not** the entropy that describes the adversary's uncertainty in a given attack scenario.

5. CONCLUSIONS

Shannon entropy has proven to be very useful to measure anonymity, both in the application (user profiles) and communication layers. Often, entropy-based anonymity metrics have been used for the evaluation of mixes, without taking into account any user profile information the attacker could have. Integrating several sources of information has not yet been fully addressed.

As Shannon proves that $H(Y|X) \leq H(Y)$, the confusion between the adversary's uncertainty (which is represented by the "entropy of a conditional probability distribution") and the "conditional entropy" has led some researchers to mistakenly believe that access to more information in any concrete case must necessarily result in a reduction of the adversary's uncertainty.

We have explained the relationship between the adversary's uncertainty, which is calculated from a particular observation, and Shannon's conditional entropy (which averages all possible cases), and shown an example where the recipient anonymity of a message increases when the adversary combines two sources of information (profiling and network observation).

Acknowledgements. This work was partially supported by the IWT SBO ADAPID project (Advanced Applications for e-ID cards in Flanders), GOA Ambiorics and IUAP p6/26 BCRYPT. George Danezis is funded by a research grant of the Katholieke Universiteit Leuven. Carmela Troncoso is funded by a grant of the Foundation Barrie De la Maza from Spain.

6. REFERENCES

[1] Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, 1(6):27–34, 2003.

[2] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.

[3] Sebastian Clauß. A framework for quantification of linkability within a privacy-enhancing identity management system. In *Emerging Trends in Information and Communication Security*, pages 191–205. Springer, LNCS 3995, 2006.

[4] Sebastian Clauß and Stefan Schiffner. Structuring anonymity metrics. *Proceedings of the ACM Workshop on Digital Identity Management*, pages 55–62, 2006.

[5] George Danezis. Statistical disclosure attacks: Traffic confirmation in open environments. In *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–426. Kluwer, 2003.

[6] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In *Information Hiding*, pages 293–308. Springer, LNCS 3200, 2004.

[7] Claudia Diaz, Joris Claessens, Stefaan Seys, and Bart Preneel. Information theory and anonymity. In B. Macq and J.-J. Quisquater, editors, *Werkgemeenschap voor Informatie en Communicatietheorie*, pages 179–186, 2002.

[8] Claudia Diaz and Bart Preneel. Reasoning about the anonymity provided by pool mixes that generate dummy traffic. In *Information Hiding*, pages 309–325. Springer, LNCS 3200, 2004.

[9] Claudia Diaz and Andrei Serjantov. Generalising mixes. In *Designing Privacy Enhancing Technologies, Proceedings of PET'03*, pages 18–31. Springer-Verlag, LNCS 2760, 2003.

[10] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Designing Privacy Enhancing Technologies, Proceedings of PET'02*, pages 54–68. Springer-Verlag, LNCS 2482, 2002.

[11] Dogan Kesdogan and Lexi Pimenidis. The hitting set attack on anonymity protocols. In *Information Hiding*, pages 326–339. Springer, LNCS 3200, 2004.

[12] Nick Mathewson and Roger Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In *Privacy Enhancing Technologies*, pages 17–34. Springer, LNCS 3424, 2004.

[13] Andreas Pfitzmann and Marit Hansen. Anonymity, unobservability and pseudonymity – a proposal for terminology. In *Designing Privacy Enhancing Technologies, Proceedings of PET'00*, pages 1–9. Springer-Verlag, LNCS 2009, 2001.

[14] Alfred Rényi. On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium Mathematical Statistics and Probability*, 1:547–561, 1961.

[15] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *Designing Privacy Enhancing Technologies, Proceedings of PET'02*, pages 41–53. Springer-Verlag, LNCS 2482, 2002.

[16] Claude Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423:623–656, 1948.

[17] Sandra Steinbrecher and Stefan Köpsell. Modelling unlinkability. In *Designing Privacy Enhancing Technologies, Proceedings of PET'03*, pages 32–47. Springer-Verlag, LNCS 2760, 2003.

[18] Gergely Tóth, Zoltán Hornák, and Ferenc Vajda. Measuring anonymity revisited. In *Proceedings of the Ninth Nordic Workshop on Secure IT Systems*, pages 85–90, 2004.