Measuring Anonymity Revisited

Gergely Tóth, Zoltán Hornák and Ferenc Vajda
Budapest University of Technology and Economics
Department of Measurement and Information Systems
H-1111 Budapest, XI., Magyar tudósok krt. 2.
Email: {tgm,hornak,vajda}@mit.bme.hu

***Abstract*— Anonymous message transmission systems are the building blocks of several high-level anonymity services (e.g. e-payment, e-voting). Therefore, it is essential to give a theoretically based but also practically usable objective numerical measure for the provided level of anonymity. In this paper two entropy-based anonymity measures will be analyzed and some shortcomings of these methods will be highlighted. Finally, source- and destination-hiding properties will be introduced for so called *local* anonymity, an aspect reflecting the point of view of the users.**

***Index Terms*— anonymity measure, local anonymity**

I. Introduction

Anonymous message sending techniques define a rapidlyn evolving area in privacy research. Such methods are required for many applications ranging from simple untracable e-mail communication to anonymous electronic voting and payment systems. The goal is to transport messages from senders to recipients, so that an attacker (ranging from simple observers to traffic shaping adversaries) can only guess the user-message relations with a small probability.

The aim of this paper is to draw attention to so called *local* anonymity. Recent papers have proposed entropy as a means of measuring the performance of different systems. Although *global* anonymity (i.e. how many potential candidates the adversary has to consider and what is their general distribution) can be quantified this way, the user's point of view is somewhat different: "I am only interested in my own messages and they should not be linked to me under any circumstances with a probability greater than a given treshhold". In response to this we should rather focus on the worst case scenario for a given message.

Another key issue is the aspect of the *user-defined treshhold*. This is a calibration metric, like Quality-of-Service that a system should satisfy when providing anonymity services. The aim in this paper is to clearly highlight the problem – the difference between local and global anonymity – and give some example solutions.

Although one can find several systems in the literature ([1], [2], [3]), each year newer and newer solutions are published ([4], [5]). In order to objectively compare them, appropriate theoretical measures are required. Another important requirement of the practical usability of such measures is that the measure should be easy to understand, not only by experts but also by users. The goal of the authors is to introduce source-hiding property for measuring sender anonymity and destination-hiding property for recipient anonymity and to compare them with existing measures.

A. Local vs. Global Anonymity

Serjantov & Danezis [6] and Díaz *et al* [7] proposed two similar information theory-based anonymity measures. By using the entropy of the attacker's probability distribution, they quantified how many bits of information an adversary needs in order to perfectly match a message to the respective user. This approach (later referred to as as *global* measure) aims to quantify the effort that is needed to totally compromize messages. (In the worst case missing bits of information can be substituted with brute force, where the required number of steps is the power of two.)

On the other hand, in this paper we argue that another approach – using the maximal probability as a measure – focuses better on the *local* aspect of anonymity. From the users' point of view this is more important, because they are interested only in their own messages and the probability of being compromized.

B. Outline of the Paper

In Section II. we briefly introduce previous work in the field of anonymity measures and then analyze the shortcomings of these approaches in Section III. In Section IV. the proposed source- and destination-hiding properties will be introduced. We will show that they represent worst-case anonymity measures, mainly focusing on the local aspect of the user's view. Finally the analysis of a continuous time system (the PROB-channel) closes the paper with the calculations for the different anonymity metrics, which have been introduced.

II. Background

This section gives a short introduction to the background of measuring anonymity. First let the informal summary of an anonymous message transmission system follow. This will define the terms and expressions we are going to use in this paper. Later in this section different previously published entropy-based anonymity measures will be described.

A. Anonymous Message-sending Scenario

In an anonymous message-sending scenario we have the following setting: *senders* send *messages* to *recipients* using the intermediate *anonymous message transmission system.* This anonymous message transmission system cyptographically *transforms*, *delays* and *mixes* the messages sent by the senders according to the implemented algorithm and eventually delivers them to the recipients.

On the other hand there is an *adversary*, who may see messages sent by the senders and also those delivered to the recipients. His aim is to match the delivered ones to the senders (accoring to [8] in this case sender anonymity is compromized) or the sent messages to the recipients (recipient anonymity).

In order to render the effors of the adversary more difficult, the parties use diffent encryption algorithms, uniformly sized messages and dummy traffic [9].

Considering the adversary different attacker models can be taken into account: mighty ones may perceive the whole network at all times, whereas a less pessimistic approach may consider attackers with limited access to a fraction of the whole network. Another important aspect is whether the adversary is active (i.e. may delay, create, delete or alter messages) or only passive (i.e. can only eavesdrop). When calculating the level of anonymity provided by a system it is an important aspect to note against what kind of adversary the metrics hold.

Furthermore we assume that the adversary performs a probabilistic attack: he computes probabilities that indicate, to what extent messages correspond to senders or recipients according to his knowledge. Finally the adversary marks the most probable sender/recipient as his guessed user for a certain message.

B. Anonymity Measures

Based on the model of an anonymous message transmission system the definition of anonymity was given by Pfitzmann and Köhntopp [8]:

> Anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*.
> [...]
> Anonymity may be defined as the unlinkability of an IOI[1] and an identifier of a subject.

The first publications aiming to quantify the level of anonymity provided by the described systems used the size of the anonymity set as the measure (e.g. [3]). Since the probabilities might not be uniformly distributed, the size of the set does not perfectly reflect the achieved anonymity as it was pointed out with the practical example of the pool mix in [10].

Based on the above observation Serjantov & Danezis introduced entropy for measuring anonymity [6]. They used the following model:

Definition 1: Given a model of the attacker and a finite set of all users Ψ, let $r \in \mathcal{R}$ be a role for a user ($\mathcal{R}$={sender, recipient}) with respect to a message $\mathcal{M}$. Let $\mathcal{U}$ bet the attacker's a-posteriori probability of users $u \in \Psi$ having the role r with respect to $\mathcal{M}$.

With this in mind the measure for both sender and recipient anonymity was defined as follows:

Definition 2: The effective size S of an r anonymity probability distribution $\mathcal{U}$ is equal to the entropy of the distribution. In other words

$$S = -\sum_{u \in \Psi} p_u \log_2 p_u \qquad (1)$$

where $p_u = \mathcal{U}(u, r)$.

In the rest of the paper this anonymity measure will be referred to as the *simple entropy* measure.

Díaz et al. followed a slightly different (extended) approach [7], whereas they only considered sender anonymity. Let $\mathcal{A}$ represent the anonymity set of a certain message $\mathcal{M}$, i.e. $\mathcal{A} = \{u | (u \in \Psi) \wedge (p_u > 0)\}$. Furthermore let N be the size of the anonymity set, i.e. $N = |\mathcal{A}|$. Their defintion was the following:

Definition 3: The degree of anonymity provided by a system is defined by

$$d = \frac{H(X)}{H_M} \qquad (2)$$

For the particular case of one user we assume d to be zero.

With the symbols defined above $H(X) = S$ and $H_M = \log_2 N$. We will refer to this measure as the *normalized entropy* measure.

In both cases 0 means absolutely no anonymity (i.e. the attacker knows with 100% the sender of a message). In the simple entropy case maximal anonymity is achieved when $S = \log_2 N$ and with normalized entropy when $d = 1$.

III. Shortcomings of Existing Anonymity Measures

In the following the previously introduced entropy based measures will be evaluated and some shortcomings will be pointed out:

- For both measures two probability distributions will be given that have the same level of anonymity according to the respective measure, but practically provide very different anonymity considering the local aspect, i.e. the worst case for one particular user.
- It will be shown that non-desirable systems can approach optimal systems according to the entropy based measures.

A. Simple Entropy

Recall that accoring to the measure of simple entropy the level of anonymity is given by S, see (1). First, two distributions will be shown that have the same entropy but behave remarkably differently considering the provided anonymity from one user's point of view.

Now let's define the following two anonymity systems:

1) In the first system the probability distribution (D_1) is uniform among m users, i.e. $D_1 : p_u = \frac{1}{m}$.
2) In the second system we have a different distribution (D_2) among n users: for the sake of the example for the

[1] Item Of Interest, i.e. a message

TABLE I

CORRESPONDING n-m VALUES YIELDING THE SAME AMOUNT OF SIMPLE ENTROPY

m	n	S
10	26	3.3219
20	101	4.3219
50	626	5.6439
100	2501	6.6439

actual sender the probability is 50% and the others are uniformly distributed [2]. This yields the following:

$$D_2\colon p_u = \begin{cases} 0.5 & \text{for the actual sender,} \\ \frac{0.5}{n-1} & \text{otherwise.} \end{cases}$$

Now let's choose m and n so that the resulting entropy is the same. For this we have to solve $S^m_{D_1} = S^n_{D_2}$, which is expanded in the following equation:

$$-\left[m\frac{1}{m}\log_2\frac{1}{m}\right] = \\ -\left[(n-1)\frac{0.5}{n-1}\log_2\frac{0.5}{n-1} + 0.5\log_2 0.5\right] \qquad (3)$$

The result can be seen in (4):

$$n = \frac{m^2}{4} + 1 \qquad (4)$$

Some example numerical values are shown in Table I. In order to visualize the problem, let's have a look at the example with $m = 20$. According to the definitions in such a system with uniformly distributed probabilities (D_1) the attacker has 5% (i.e. $p_u = \frac{1}{20} = 0.05$) chance to guess the sender of a message. This system provides anonymity with $S = 4.3219$ bits.

On the other hand, let's have a look at the second system (D_2). Here for each delivered message the attacker knows that a particular sender sent the message with 50% certainty and another 100 senders could have sent it with 0.5%.

The two systems clearly perform differently considering the local aspect of anonymity, but have the same value with simple entropy. With D_1 distribution statistically seen on the long term an attacker can guess the sender of a message every 20th time correctly, whereas with D_2 distribution he is going to successfully guess the sender of every second message.

The second point to show is that non-desirable systems can achieve an arbitrarily high entropy. From (1) and (4) it is clear that for an arbitrary value of S a corresponding D_2 distribution can be constructed, where $n = 4^{S-1} + 1$.

Summarized, the main problem with this entropy based measure is that it tries to quantify the amount of information that is required to break *totally* the anonymity of a message, i.e. to definitely identify actual sender of a message. However in practice we have to consider an attacker successful, if he can guess the sender of some selected messages with a good probability, in specific cases significantly greater than in the case of the uniform distribution (i.e. $p_u \gg \frac{1}{N}$).

B. Normalized Entropy

To demonstrate similar shortcomings of the normalized entropy measure first we show two systems with the same value of d, however with remarkably different local anonymity. Due to the normalization we have to notice that following from the definition of d in order to obtain the same results for the two constructions the quotient of the entropy and the the logarithm of the anonymity set size should remain the same. This can be achieved in the easiest way by having the same entropy as well as the same anonymity set size.

For demonstration purposes let's consider the following two systems:

1) In the first system we have the distribution (D_2) known from the previous example: n users are involved; for the actual sender the probability is 50% and the others are uniformly distributed. This yields the following distribution:

$$D_2\colon p_u = \begin{cases} 0.5 & \text{for the actual sender,} \\ \frac{0.5}{n-1} & \text{otherwise.} \end{cases}$$

2) In the second case we have a new distribution (D_3): there are n users as well, x of them having a probability P_A and $n - x$ of them with probability P_B. The characteristic parameters for such a distribution are x and P_S, being the sum of the P_A probabilities of the x users. The following distribution is given this way:

$$D_3\colon p_u = \begin{cases} P_A = \frac{P_S}{x} & \text{for the } x \text{ users,} \\ P_B = \frac{1-P_S}{n-x} & \text{otherwise.} \end{cases}$$

This second distribution can be explained as follows: with P_S probability the sender of the message is the member of a sub-anonymity-set $\mathcal{A}_A$ with x members and uniformly distributed probabilities P_A. On the other hand with $1-P_S$ probability the sender of the message is the member of the other sub-anonymity-set $\mathcal{A}_B$ with $n-x$ members and also uniformly distributed probabilities P_B.

In order to find suitable distributions the equation below has to be solved (notice that for the distribution D_2 the entropy was caluclated in (3)):

$$\frac{-\left[(n-1)\frac{0.5}{n-1}\log_2\frac{0.5}{n-1} + 0.5\log_2 0.5\right]}{\log_2 n} = \\ \frac{-\left[xP_A\log_2 P_A + (n-x)P_B\log_2 P_B\right]}{-log_2 n} \qquad (5)$$

It is clear that for this scenario we have three variables: n, x and P_S. For a concrete example, x was chosed to be $x = \frac{m}{2}$, where $m = \sqrt{4n-4}$ (see (4)) and the respective P_S was calculated (see Table II).

To imagine the two systems, let's look at the case $m = 20$. With this we get an anonymity set $\mathcal{A}$ with $n = 101$ users. For

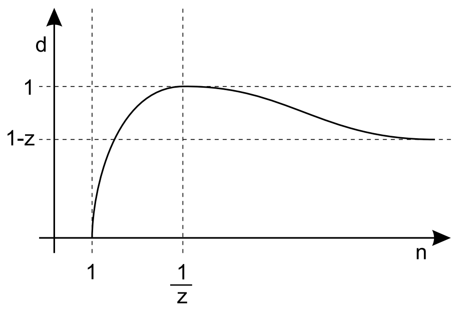
[2] The concrete probability of 0.5 was chosen in order to simplify the resulting equations.

TABLE II

CORRESPONDING n-x-P_S VALUES YIELDING THE SAME NORMALIZED ENTROPY

m	n	x	P_S	P_A	P_B	d
10	25	5	0.832213	0.166442	0.007989	0.706727
20	101	10	0.865184	0.086518	0.001481	0.649112
50	626	25	0.890594	0.035623	0.000182	0.607518
100	2501	50	0.903463	0.018069	0.000039	0.588561

Fig. 1. d as a function of n with distribution D'_2



both systems the normalized entropy gives $d = 0.649112$ as a measure for the anonymity.

In case of the first system (D_2) for the actual sender of the message $p_u = 0.5$, thus the attacker knows with 50% certainty of the sender, for the other 50% he has 100 possible users with 0.5% certainty uniformly distributed.

On the other hand for the second system (D_3) we have two sub-anonymity-sets. For $\mathcal{A}_A$ we have 10 users with probabilities P_A of roughly 8.7%, yielding together P_S of 87%. Furthermore we have the other sub-anonymity-set $\mathcal{A}_B$, consisting of 91 users with an overall probability of about 13% uniformly distributed in quantities of 0.15% as P_B.

Another important point is to show that non-desirable systems exist in arbitrarily small vicinity of the optimal $d = 1$. For this let's consider a slightly modified version of the D_2 distribution that we will refer to as D'_2. In this distribution n users are involved; for the actual sender the probability is z and the others are uniformly distributed. This yields the following:

$$D'_2 \colon p_u = \begin{cases} z & \text{for the actual sender,} \\ \frac{1-z}{n-1} & \text{otherwise.} \end{cases}$$

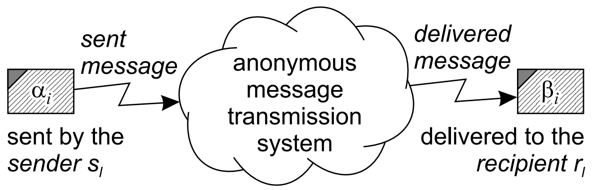
The degree of anonymity provided according to the normalized entropy is as follows:

$$d = \frac{-\left(z \log_2 z + (n-1)\frac{1-z}{n-1} \log_2 \frac{1-z}{n-1}\right)}{\log_2 n} \qquad (6)$$

After analyzing d as a function of n (as seen on Fig. 1) we can determine the following:

- With one user $d = 0$ is trivial.
- With $\frac{1}{z}$ users $d = 1$ is maximal. This is evident as in this case we have uniform distribution.
- Finally it can be proven that $\lim_{n\to\infty} d = 1 - z$ and that on the interval $\left(\frac{1}{z}, \infty\right)$ $d > 1 - z$.

Fig. 2. Message sending with the anonymous message transmission system



With the above in mind we can see that even with a system, where $n \gg \frac{1}{z}$ the degree of anonymity is above the treshhold, i.e. $d > 1 - z$, thus systems can get arbitrarily close to the optimal $d = 1$ and yet they are non-desirable in the sense that there are users whose level of local anonymity is above an acceptable probability.

IV. LOCAL ANONYMITY MEASURE

In the previous section shortcomings of the information theory based *global* anonymity metrics were evaluated. In those cases it was quantified, how much additional information an attacker needs in order to definitely identify the user corresponding to the message (i.e. its sender or recipient).

On the contrary our argument is that an attacker is already successful if he can guess these links with a good probability (that is over a certain acceptable treshhold).

Before defining local anonymity measures, the used terms will be introduced. In the analyzed system *senders* ($s_l \in S$) transmit encrypted *sent messages* ($\alpha_j \in \varepsilon_S$) to the anonymous transmission system. After transforming (re-encoding) and delaying them the *delivered messages* ($\beta_k \in \varepsilon_R$) reach the *recipients* (see Fig. 2.). Time of sending is indicated by $t_S(\alpha_j)$, similarly time of receipt is $t_R(\beta_k)$. Sender of a sent message is denoted by $S(\alpha_j)$ and the recipient by $R(\beta_k)$.

The adversary has two aims: to break sender anonymity by computing the probabilities P_{β_k,s_l} (i.e. what is the probability that β_k was sent by s_l) and to break recipient anonymity by computing P_{α_j,r_l} (i.e. r_l received α_j).

For this scenario in [5] the destination- and source-hiding properties were defined for sender and recipient *local* anonymity.

Definition 4: A system is *source-hiding* with parameter Θ if the adversary cannot assign a sender to a delivered message with a probability greater than Θ, i.e. if

$$\forall_{\beta_k} \forall_{s_l} \left(P_{\beta_k,s_l} \leq \Theta\right) \qquad (7)$$

holds.

Definition 5: A system is *destination-hiding* with parameter Ω if the adversary cannot assign a recipient to a sent message with a probability greater than Ω, i.e. if

$$\forall_{\alpha_j} \forall_{r_l} \left(P_{\alpha_j,r_l} \leq \Omega\right) \qquad (8)$$

holds.

It is important to note that one cannot draw grounded conclusions about the local anonymity from global anonymity measures as it was shown in the previous section (i.e. for arbitrarily high global anonymity systems with non-desirable local anonymity exist, where in the worst case the identity of some users can be guessed with an unacceptably big probability).

On the contrary we will show that from the local anonymity measures we can draw conclusions for the global anonymity meausres as well. In the following we will deduce results for the sender anonymity from the source-hiding property but since it is symmetric for the destination-hiding property, similar equations can be stated as well for the recipient anonymity.

Theorem 1: For a system with source hiding property with parameter Θ the inequality below holds:

$$S \geq -\log_2 \Theta \tag{9}$$

Informally this theorem means that a system of source-hiding property with parameter Θ is in the global sense at least as strong as a system with $\frac{1}{\Theta}$ users and uniformly distrbuted probabilities.

Proof: First from the definition (7) it follows that $\forall_{u \in \Psi}(0 < p_u \leq \Theta \leq 1)$. Therefore since the logarithm function is monotonic in the interval $(0, \infty)$ $\Rightarrow$ $\forall_{u \in \Psi}(\log_2 p_u \leq \log_2 \Theta) \Rightarrow \forall_{u \in \Psi}(-\log_2 p_u \geq -\log_2 \Theta)$.

With this (1) can be rewritten:

$$\begin{aligned} S &= -\sum_{u \in \Psi} p_u \log_2 p_u \\ &\geq -\sum_{u \in \Psi} p_u \log_2 \Theta \\ &= -\log_2 \Theta \sum_{u \in \Psi} p_u \\ &= -\log_2 \Theta. \end{aligned}$$

■

With the combination of (2) and (9) a lower limit to d can be given as well:

$$d \geq -\log_N \Theta \tag{10}$$

as $d = \frac{S}{\log_2 N} \geq -\frac{\log_2 \Theta}{\log_2 N} = -\log_N \Theta$.

V. Analysis of the PROB-channel

The PROB-channel is an example for an anonymous message transmission system introduced in [5]. In order to show how the introduced anonymity metrics work with practical anonymity systems in this section the PROB-channel will be introduced and its anonymioty level will be analyzed.

A. Brief Defintion of the PROB-channel

The PROB-channel is a continuous time system, where messages are processed independently. Once a message enters the channel, a delay will be calculated for it and after that time has passed the message leaves. This delay δ in the system is a probability variable with a given density function $f(\delta)$ (i.e. $\int_0^\infty f(\delta) d\delta = 1$). In order to guarantee real-time probabilities, the delay in the PROB-channel has a pre-defined maximum (δ_{max}). On the other hand considering real systems a minimal delay (δ_{min}) was also defined:

$$\forall_{\delta \notin (\delta_{\text{min}}, \delta_{\text{max}})} f(\delta) = 0 \tag{11}$$

In order to simplify further equations first two sets need to be defined. With μ_{β_k} the set of sent messages is meant that might have left the channel as β_k (12), whereas η_{β_k, s_l} denotes the subset of μ_{β_k}, which was sent by the specific sender s_l (13).

$$\begin{aligned} \mu_{\beta_k} = \{\alpha_j | \, (t_R(\beta_k) - \delta_{\text{max}}) < t_S(\alpha_j) < \\ (t_R(\beta_k) - \delta_{\text{min}})\} \end{aligned} \tag{12}$$

$$\eta_{\beta_k, s_l} = \{\alpha_j | (\alpha_j \in \mu_{\beta_k}) \wedge (S(\alpha_j) = s_l)\} \tag{13}$$

B. The Attacker Model – A Passive Observer

As an attacker model let's consider a passive observer: he can eavesdrop on all connections but does not alter, delete or delay messgaes.

Aim of the passive observer is to link delivered messages to the senders by computing the probabilities P_{β_k, s_l}. The most effective solution is summarized in (14):

$$P_{\beta_k, s_l} = \frac{\sum_{\alpha_j \in \eta_{\beta_k, s_l}} f\left(t_R(\beta_k) - t_S(\alpha_j)\right)}{\sum_{\alpha_j \in \mu_{\beta_k}} f\left(t_R(\beta_k) - t_S(\alpha_j)\right)} \tag{14}$$

Of course the attacker chooses s_i as the sender for β_k where $s_i = \max_{s_l} P_{\beta_k, s_l}$.

C. Methods to Ensure Local Anonymity

It it clear from (14) that in the general case no hard guarantee can be given about P_{β_k, s_l}. The main problem comes from the real-time requirement: even if only one message is in the channel, it has to be delivered before the maximal delay expires. Thus in unfortunate cases the adversary has an easy task.

In order to ensure that there are enough messages to form a sufficiently large anonymity set for each message the only solution is to enforce continuous message sending. The MIX/MAX property was defined for this purpose.

Definition 6: A system fulfills the criteria of the MIX/MAX property with parameters τ_{min} and τ_{max} ($\tau_{\text{min}} < \tau_{\text{max}} < \delta_{\text{max}}$) if all senders send at least one message in every τ_{max} interval but no sender sends more than one message in any τ_{min} interval.

With the above definiton the amount of messages can be fine-tuned and also the fraction, a specific user reaches from the whole amount of messages, can be set. It was shown in [5] that with the MIX/MAX property local anonymity can be ensured, see (15) for the source-hiding property.

$$\Theta = \frac{\sum_{i=1}^{\Delta_{\text{min}}} \max_{(i-1) \cdot \tau_{\text{min}} \leq q \leq i \cdot \tau_{\text{min}}} f(q)}{N \cdot \sum_{i=1}^{\Delta_{\text{max}}} \min_{(i-1) \cdot \tau_{\text{max}} \leq q \leq i \cdot \tau_{\text{max}}} f(q)} \tag{15}$$

where $\Delta_{\text{max}} = \lfloor \frac{\delta_{\text{max}} - \delta_{\text{min}}}{\tau_{\text{max}}} \rfloor$ and $\Delta_{\text{min}} = \lceil \frac{\delta_{\text{max}} - \delta_{\text{min}}}{\tau_{\text{min}}} \rceil$.

D. The Optimal System

Sticking to the real-time guarantee if was proven that the optimal delay characteristics is achieved if the channel uses the uniform density function (16).

$$f(\delta) = \frac{1}{\delta_{\text{max}} - \delta_{\text{min}}} \tag{16}$$

It is interesting to note that by changing the guaranteed maximal delay to a softer mean delay of a and enabling even infinite delays (of course with small probability) another density function proves to be the optimal, namely the exponential one (17) as first proposed by Kesdogan et al. for the SG-MIX [11] and then proven to be optimal by Danezis [12].

$$f(\delta) = \frac{1}{a} e^{-\frac{1}{a}\delta} \tag{17}$$

E. Quantified Anonymity of the PROB-channel

In the optimal case of uniform delay density (16) and MIX/MAX property (15) the local anonymity can be guaranteed efficiently. If $N \geq \frac{\tau_{\text{max}}}{\tau_{\text{min}}}$ then the following equation gives a good approximation:

$$\Theta \approx \frac{\tau_{\text{max}}}{N \cdot \tau_{\text{min}}} \tag{18}$$

Using results (9) and (10) from the previous section the following guarantees can be given for the global anonymity:

$$\begin{aligned} S &\geq -\log_2 \Theta \\ &= \log_2 N - \log_2 \frac{\tau_{\text{max}}}{\tau_{\text{min}}} \end{aligned} \tag{19}$$

$$\begin{aligned} d &\geq -\log_N \Theta \\ &= 1 - \log_N \frac{\tau_{\text{max}}}{\tau_{\text{min}}} \end{aligned} \tag{20}$$

From these results it is clear that the main calibration possibility of the PROB-channel is the fraction $\frac{\tau_{\text{max}}}{\tau_{\text{min}}}$. It is obvious that if $\tau_{\text{max}} = \tau_{\text{min}}$ then an overall optimum can be reached where the anonymity set of each message is maximal and the probabilities P_{β_k, s_l} are uniformly distributed among all possible senders:

$$P_{\beta_k, s_l} = \frac{1}{N} \tag{21}$$

VI. Conclusion

The main focus of this paper was to introduce the term *local* anonymity and appropriate metrics for measuring it: the source- and destination-hiding properties. Previous information theory based anonymity measures aimed to quantify the number of bits required by an adversary to completely trace back a message. On the contrary we argue that an attacker is already successful if he can compromize messages with a probability above a certain treshhold for some of the users – which from the local aspect of the users is unacceptable, however possible in unfortunate cases of entropy-based global anonymity measures.

With this paper the importance of the local aspect was underlined and via a practical example of the PROB-channel enlightend. Future work should be carried out in order to analyze other practical solutions as well from this local point of view.

References

[1] D. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Communications of the ACM*, vol. 4, no. 2, pp. 84–88, February 1981.

[2] M. Reiter and A. Rubin, “Crowds: Anonymity for web transactions,” *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66–92, June 1998.

[3] O. Berthold, H. Federrath, and S. Köpsell, “Web mixes: A system for anonymous and unobservable internet access,” in *Designing Privacy Enhancing Technologies*, ser. Springer-Verlag, LNCS, H. Federrath, Ed., vol. 2009, Berkeley, CA, 2001, pp. 115–129.

[4] R. Dingledine, N. Mathewson, and P. Syverson, “Tor: The second-generation onion router,” in *Proceedings of the 13th USENIX Security Symposium*, August 2004.

[5] G. Tóth and Z. Hornák, “Measuring anonymity in a non-adaptive, real-time system,” in *Proceedings of Privacy Enhancing Technologies (PET2004)*, ser. Springer-Verlag, LNCS, Forthcoming, 2004.

[6] A. Serjantov and G. Danezis, “Towards an information theoretic metric for anonymity,” in *Proceedings of Privacy Enhancing Technologies (PET2002)*, ser. Springer-Verlag, LNCS, P. Syverson and R. Dingledine, Eds., vol. 2482, San Francisco, CA, April 2002.

[7] C. Díaz, S. Seys, J. Claessens, and B. Preneel, “Towards measuring anonymity,” in *Proceedings of Privacy Enhancing Technologies (PET2002)*, ser. Springer-Verlag, LNCS, P. Syverson and R. Dingledine, Eds., vol. 2482, San Francisco, CA, April 2002, pp. 54–68.

[8] A. Pfitzmann and M. Köhntopp, “Anonymity, unobservability, and pseudonymity – a proposal for terminology,” in *Designing Privacy Enhancing Technologies*, ser. Springer-Verlag, LNCS, H. Federrath, Ed., vol. 2009, Berkeley, CA, 2001, pp. 1–9.

[9] O. Berthold and H. Langos, “Dummy traffic against long term intersection attacks,” in *Proceedings of Privacy Enhancing Technologies workshop (PET 2002)*, R. Dingledine and P. Syverson, Eds., vol. 2482. San Francisco, CA: Springer-Verlag, LNCS, April 2002.

[10] A. Serjantov and R. E. Newman, “On the anonymity of timed pool mixes,” in *Security and Privacy in the Age of Uncertainty*. Athens, Greece: Kluwer, May 2003, pp. 427–434, (Workshop on Privacy and Anonymity Issues in Networked and Distributed Systems).

[11] D. Kesdogan, J. Egner, and R. Büschkes, “Stop-and-go MIXes: Providing probabilistic anonymity in an open system,” in *Proceedings of Information Hiding Workshop (IH 1998)*, ser. Springer-Verlag, LNCS, vol. 1525, Berkeley, CA, 1998.

[12] G. Danezis, “The traffic analysis of continuous-time mixes,” in *Proceedings of Privacy Enhancing Technologies (PET2004)*, ser. Springer-Verlag, LNCS, Forthcoming, 2004.