Recipient Privacy in Online Social Networks

Filipe Beato[1], Kimmo Halunen[2], and Bart Mennink[1]

[1] Dept. Electrical Engineering, ESAT/COSIC, KU Leuven, and iMinds, Belgium
{firstname.lastname}@esat.kuleuven.be
[2] VTT Technical Research Centre of Finland Ltd, Oulu, Finland
kimmo.halunen@vtt.fi

Abstract. Alongside the intensive growth of Online Social Networks (OSNs), privacy has become an important concept and requirement when sharing content online, leading users to enforce privacy often using encryption when sharing content with multiple recipients. Although cryptographic systems achieve common privacy goals such as confidentiality, key privacy, and recipient privacy, they have not been designed aiming at dynamic types of networks. In fact, the interactive nature of OSNs provides adversaries new attack vectors against privacy, and in particular against recipient privacy. We present the notion of frientropy, and argue that privacy of recipients is maintained in OSNs provided that the social graph has a high frientropy, besides the conventional recipient privacy notion. We compute the frientropy for various theoretical settings, and discuss its implications on some practical settings.

Keywords: recipient privacy; entropy; online social networks

1 Introduction

In today's networked and interactive world privacy has become a focal point of research, and it is debated over several disciplines, such as law, philosophy, technology, and cryptography. Privacy-enhancing technologies have been developed and both lauded and critiqued by experts, politicians, and other interested parties. Although finding a suitable definition for privacy is a very hard task, in modern cryptography the requirement of privacy is resumed to provable security and privacy definitions. In general, notions such as k-anonymity [16] and anonymity sets [14] have been used to make explicit mathematical formulations on anonymity. In the public key cryptography setting, Bellare *et al.* defined the notion of public key anonymity as *key privacy* [6], so that an adversary cannot distinguish which key was used for the encryption and thus cannot identify the user from the ciphertext. Later, Barth *et al.* [2] extended this notion for the multiple user setting by *recipient privacy*, such that to achieve privacy of the recipients on broadcast encryption, the adversary cannot distinguish the identities in the authorized recipient set. The concept has been generalized by Libert *et al.* [13], and relaxed by Fazio *et al.* [11] with the outsider recipient privacy notion, towards outsiders and not against authorized users, i.e., insiders.

At the same time, with the enormous growth of interactive services such as Online Social Networks (OSNs), different privacy paradigms have been introduced, allowing users to share content with multiple entities which may interact with replies, likes, and comments. Yet, the above notions of recipient privacy have not been designed for dynamic systems, and become void and difficult to maintain, as each reply gives away the identity of at least one of the recipients, and subsequently compromises the privacy of the whole recipient set in the adversarial model. In fact, current definitions are designed for scenarios where the data is conveyed through some content distribution system and accessed by different users, and not published on interactive centralized platforms susceptible to single replies revealing the users in the set. Hence, current definitions assume that adversaries are not able to gain information of the set by observing interactions in the communications within the set, which is unrealistic for OSNs.

In this paper, we take a look at the problem of *recipient privacy* when applied to interactive platforms, used in cases where the aim of the cryptographic scheme is to enable anonymous and confidential communications between several entities over a potentially malicious platform with interactive properties, such as OSNs. This notion has been used by practical tools, such as Scramble! [5], aiming at protecting the privacy and secrecy of the content shared in OSNs [3, 4]. We begin by demonstrating the shortcomings (Section 3) of the existing recipient privacy definitions when applied to interactive platforms such as OSNs, as the security and privacy assumptions as well as the proofs of previous constructions can be broken in a straightforward fashion. Hence, in order to re-define the notion of recipient privacy with interactive properties we introduce the novel notion of *frientropy* (Section 4), apply it to complete and threshold sets of recipients and discuss the practical impact of frientropy (Section 5). Then we present a new interactive recipient privacy definition along with the notion of membership queries (Section 6). We argue that the new notion presents a more suitable definition when applied to interactive platforms and that solely the encryption of the message content does not suffice to provide privacy of recipients of messages against realistic adversaries in the OSN setting. Finally, we discuss our results in the new setting and conclude with future research topics.

2 Private Broadcast Encryption

For any $n \in \mathbb{N}$, let $\{0,1\}^n$ denote the set of bit strings of length n, and $\{0,1\}^*$ the set of bit strings of arbitrary length. For two strings x and y, $x \parallel y$ denotes their concatenation and $x \oplus y$ their bitwise XOR. The notation $x \overset{\$}{\leftarrow} X$ indicates that x is selected uniformly at random from the finite set X and $x \overset{\mathcal{X}}{\leftarrow} X$ that x is selected from X according to some arbitrary probability distribution $\mathcal{X}$ over X. For any two sets X and Y, we define the union by $X \cup Y = \{z : z \in X \vee z \in Y\}$, the intersection as $X \cap Y = \{z : z \in X \wedge z \in Y\}$, and the empty set by $\emptyset$.

In broadcast encryption, a sender broadcasts an encrypted message to multiple people, so that only a pre-selected set $\mathcal{S} \subseteq \mathcal{U}$ is able to decrypt, where $\mathcal{U}$ is the universe (e.g., the set of all people being active on an OSN). The goal of private broadcast encryption is to keep the pre-selected set of recipients $\mathcal{S}$ private, such that from the ciphertext it should be impossible to infer anything about the intended recipients. The security of private broadcast encryption is conventionally split into key privacy and recipient privacy.

Public key anonymity, or key privacy, was defined by Bellare *et al.* [6] as the indistinguishability property of the public keys used for encryption.

Definition 1 (Key Privacy, Bellare *et al.* [6]). *A public key encryption scheme* $\mathtt{C} \leftarrow \mathtt{Enc}_{pk}(\mathtt{m})$*, is key private if any bounded adversary* $\mathcal{A}$*, with access to the list of public keys* $\{pk_1,\ldots,pk_n\}$*, is not able to distinguish the output* $\mathtt{C}_b$ *of* $\mathtt{Enc}_{pk_b}(\mathtt{m})$ *when using some* pk_x *and* pk_y*, s.t.,* $b \in \{x, y\}$ *and* $x, y \in \{1, \ldots, n\}$*, with non-negligible probability:*

$$|\mathtt{Pr}[\mathcal{A}(pk_x, pk_y, \mathtt{m}, \mathtt{C}_x) = 1] - \mathtt{Pr}[\mathcal{A}(pk_x, pk_y, \mathtt{m}, \mathtt{C}_y) = 1]| \leq \epsilon.$$

Recipient privacy was defined by Barth *et al.* [2], and defines for any two recipient sets $\mathcal{S}_0$ and $\mathcal{S}_1$ the hardness for an $\mathcal{A}$ to distinguish between a ciphertext for the recipient set $\mathcal{S}_0$, and $\mathcal{S}_1$, given that $|\mathcal{S}_0| = |\mathcal{S}_1|$ and that $\mathcal{A}$ does not possess the secret keys of any users in $\mathcal{U} \setminus (\mathcal{S}_0 \cap \mathcal{S}_1)$. A generalization to anonymous broadcast encryption (and where the challenge message may be different for both sets, if $|\mathcal{S}_0 \cap \mathcal{S}_1| = 0$) is given by Libert *et al.* [13].

Definition 2 (Recipient Privacy, Barth *et al.* [2]). *A broadcast encryption scheme* $\mathit{\Pi} \leftarrow \{\mathtt{Setup}, \mathtt{KeyGen}, \mathtt{Encrypt}, \mathtt{Decrypt}\}$ *provides recipient privacy if a PPT adversary* $\mathcal{A}$ *wins the following game with the challenger* $\mathtt{Ch}$*, only with negligible probability:*

Init: $\mathtt{Ch}$ *runs* $\mathtt{params} \leftarrow \mathtt{Setup}(\lambda)$*, and gives* $\mathcal{A}$ *the resulting* $\mathtt{params}$*.* $\mathcal{A}$ *outputs* $\mathcal{S}_0, \mathcal{S}_1 \subseteq \mathcal{U}$*, such that* $|\mathcal{S}_0| = |\mathcal{S}_1|$*.*

Setup: $\mathtt{Ch}$ *generates keys* $(pk_i, sk_i) \leftarrow \mathtt{KeyGen}(\mathtt{params}, i)$ *for each recipient* $i \in \mathcal{U}$*, and sends* pk_i *for each* $i \in \mathcal{U}$ *and* sk_i *for each* $i \in \mathcal{S}_0 \cap \mathcal{S}_1$ *to* $\mathcal{A}$*.*

Phase 1: $\mathcal{A}$ *adaptively issues decryption queries* $q_1 = (\mathtt{C}, i)$ *for any* $i \in \mathcal{U}$*, and* $\mathtt{Ch}$ *returns* $\mathtt{Decrypt}(\mathtt{params}, sk_i, \mathtt{C})$*.*

Challenge: $\mathcal{A}$ *gives* $\mathtt{Ch}$ *a message* $\mathtt{m}$*. The* $\mathtt{Ch}$ *picks a random bit* $b \in \{0, 1\}$ *and runs* $\mathtt{C}' \leftarrow \mathtt{Encrypt}(\mathtt{params}, \{i | i \in \mathcal{S}_b\}, \mathtt{m})$*, and sends* $\mathtt{C}'$ *to* $\mathcal{A}$*.*

Phase 2: $\mathcal{A}$ *adaptively issues decryption queries* $q_2 = (\mathtt{C}, i)$*, s.t.,* $\mathtt{C} \neq \mathtt{C}'$*.*

Guess: $\mathcal{A}$ *outputs a guess* $b' \in \{0, 1\}$*.*

The adversary wins if $b = b'$*, and we define* $\mathsf{Adv}_{\mathcal{A}}^{\mathrm{RecPriv}} = \left|\mathtt{Pr}[b = b'] - \frac{1}{2}\right|$*.*

3 Practical Shortcomings of Recipient Privacy

The common definition of recipient privacy was designed to protect the identities of the recipients in content sharing systems, such as shared filesystems. Nowadays, OSNs have become very popular media for sharing content and distributing information. One key element of the OSNs is that they collect and process all the data that the users post on the OSN, thus OSNs cannot be considered a trusted third party in the communications and should be incorporated into the adversarial model. Hence, posting information encrypted to enhance privacy is a valid strategy and can be done, for example, with the help of Scramble! [5]. However, the above notion of recipient privacy does not stand against the OSN and the dynamic nature of the OSNs. Let $\Pi \leftarrow \{\texttt{Setup}, \texttt{KeyGen}, \texttt{Encrypt}, \texttt{Decrypt}\}$ be a recipient private broadcast encryption scheme and $\mathcal{S}$ and $\mathcal{S}'$ be two sets of identities with $|\mathcal{S}| = |\mathcal{S}'|$ with corresponding (sets of) keys. Now, the encryption of some message $\texttt{m}$ is done with either of these sets and the adversary needs to distinguish between $\texttt{Encrypt}_{\mathcal{S}}(\texttt{m})$ and $\texttt{Encrypt}_{\mathcal{S}'}(\texttt{m})$. Recipient privacy guarantees that the adversary has only a negligible probability for achieving this by performing decryption queries to chosen ciphertexts. However, if the adversary is able to observe subsequent transactions between the communicating parties in the communications channel or OSN, e.g., likes or comments, the adversary can easily break the privacy of the recipients under the recipient privacy definition.

For example, if the adversary chooses two sets of recipients such that $\mathcal{S} \cap \mathcal{S}' = \emptyset$, then a single like, comment or reply to the message compromises the recipient privacy property for the *whole set of recipients*. That is, a single person in the set can (inadvertently) compromise the privacy for the whole set. If the user $u \in \mathcal{S}$ comments on the message, the adversary can see this and knows immediately which set ($\mathcal{S}$ or $\mathcal{S}$') was used to encrypt and thus also the identities of the remaining members of the set. This type of observation is usually not even limited to the OSN provider, but can be done also by other members of the OSN or even publicly, e.g., Twitter. Thus the notion of recipient privacy can lead to a false sense of privacy in OSNs, where at least the OSN and quite possibly also other potential adversaries can observe such interactions.

In fact, even if $\mathcal{S} \cap \mathcal{S}' \neq \emptyset$ and even if there are more than two sets from which the adversary needs to distinguish, any comment from a person that is member of only a single set breaks recipient privacy for the other members of the set. Even if the revealed identity is included in many different sets, those sets that the identity is *not in* can be excluded as possibilities and thus distinguishing becomes easier for the adversary. After a few comments from different identities, the adversary can probably make a good guess on the set of recipients. Avoiding statistical inference from the pattern of communications is very difficult and the very facts that there a) exists a lot of interaction between people in social networks, and b) this interaction is usually observable, make recipient privacy fairly hard to accomplish in real systems.

It is also worth noting that even strong schemes, such as the ones presented in [12], cannot protect against this type of inference. The anonymity and unlinkability properties shown in [12] protect against an adversary that gains access to the credentials, but not against an adversary that observes interaction between the profile content and possible recipients. Thus, the adversary can use statistical inference on the social graph to attack the privacy of users.

Furthermore, it can be argued that even in the original intended use cases of content distribution systems, there are many ways in which the adversary could learn the actual recipients of the broadcast messages. The adversary could gain access to the system logs that may contain information on who has gotten hold of the message and whether or not they have been successful in decrypting it. Of course, this requires much more effort than in the OSN setting, but ultimately it leads to the same situation that the adversary learns some members of the recipient set and can try to infer the others from social graphs such as organizational charts, and others.

More generally, encrypting the content of the replies and comments does not protect against this type of privacy attack. As the observation is based on the metadata (i.e. who is communicating with whom etc.) of the communications, the adversary does not need access to the actual content. On the other hand, this attack does not reveal anything about the content of the messages either, so the secrecy of the content is protected even if the privacy of the recipients is violated. Consequently, these observations do not highlight issues in the conventional security definitions (such as the one on recipient privacy), but rather, identify a need for a formalism of the amount of entropy provided by the set of recipients. We will address this issue in the following sections.

4 Frientropy

To take into account the interactive nature of OSNs as content sharing platforms, we argue that an additional security property is needed on top of traditional recipient privacy. At a high level, this security property, which we dub *frientropy*, captures the case that your pool of friends has a sufficiently high level of entropy. We first need a definition of (conditional) min-entropy [1].

Definition 3 ((Conditional) Min-Entropy). *Let X be a random variable. The min-entropy of X is defined as $H_\infty(X) = -\log_2(\max_x \mathtt{Pr}[X = x])$. We define the conditional min-entropy of X based on Y as $H_\infty(X \mid Y) = -\log_2(\max_x \mathtt{Pr}[X = x \mid Y])$, where $\log_2(0) := -\infty$ and $H_\infty(X \mid \mathrm{false}) := \infty$ by default.*

We are now ready to define "frientropy." Recall that $\mathcal{U}$ defines the universe. Denote its power set by $\mathbb{P}(\mathcal{U})$, and let $\mathbb{B} \subseteq \mathbb{P}(\mathcal{U}) \setminus \{\emptyset\}$ be a set of subsets of $\mathcal{U}$ such that $\emptyset \notin \mathbb{B}$. For $i = 1, \ldots, |\mathcal{U}|$, write $\mathbb{B}_i \subseteq \mathbb{B}$ to be the set of subsets of size exactly i. Frientropy informally captures the amount of randomness a set of subsets $\mathbb{B}$ offers even if some users disclosed themselves. Note that, in practical settings, $\mathbb{B}$ corresponds to all possible friend groups in an OSN. As such, the sets of recipients are heavily dependent on the communication patterns and social graphs of users. The randomness of a set $\mathcal{B} \in \mathbb{B}$ to appear will be captured by an arbitrary probability distribution $\mathcal{X}$.

Definition 4 (Frientropy). *Let $\mathcal{X}$ be a probability distribution over the power set of $\mathbb{B}$. Let $0 \leq \kappa \leq |\mathcal{U}|$, and let $\mathcal{R} \subseteq \mathcal{U}$ be a set of size $|\mathcal{R}| = \kappa$. Let $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_{|\mathcal{U}|}) \geq (0, \ldots, 0)$. The set $\mathbb{B}$ has κ-frientropy $\boldsymbol{\Phi}$ if for each $i = 1, \ldots, |\mathcal{U}|$,*

$$H_\infty(\mathcal{B} \mid \mathcal{R} \subseteq \mathcal{B}) \geq \Phi_i \,, \tag{1}$$

where $\mathcal{B} \xleftarrow{\mathcal{X}} \mathbb{B}_i$.

The definition is intuitively captured as follows: we consider the case where $\mathcal{R}$ is a set of users that revealed themselves as being part of a certain unknown set, and the frientropy of $\mathbb{B}$ than captures the amount of randomness offered by all sets that include $\mathcal{R}$. Note that a separation of the frientropy into the different sizes of the sets in $\mathbb{B}$ makes sense, given that in broadcast encryption the size of the set of recipients is known.

One should also note that frientropy is of a dual nature to the measures of anonymity presented by Diaz *et al.* [9] and extended by Serjantov and Danezis [15] and Tóth *et al.* [17]. These measures are centered around the anonymity of a sender or (usually a single) recipient of a message in some mix network that tries to hide the identity of the senders and recipients from observers and provide a metric for the anonymity of recipients. Frientropy provides a metric for the people associated with some known recipients of a message and is thus different from the previous measures.

4.1 Frientropy for Complete Sets

As a simple example, we will compute the frientropy in case $\mathbb{B}$ is a complete set, and the probability distribution is the uniform distribution: $\mathcal{B} \overset{\$}{\leftarrow} \mathbb{B}$.

Proposition 1. *Consider $\mathbb{B} = \mathbb{P}(\mathcal{U}) \setminus \{\emptyset\}$ to be the set of all non-empty subsets of $\mathcal{U}$, and let $\mathcal{X}$ be the uniform distribution. Let $0 \leq \kappa \leq |\mathcal{U}|$. The set $\mathbb{B}$ has κ-frientropy $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_{|\mathcal{U}|})$, where*

$$\Phi_i = \begin{cases} \infty \ \textit{if}\ 1 \leq i < \kappa\,, \\ \log_2\left(\binom{|\mathcal{U}|-\kappa}{i-\kappa}\right) \ \textit{if}\ i \geq \kappa\,. \end{cases}$$

Proof. Let $\mathcal{R} \subseteq \mathcal{U}$ be a set of size $|\mathcal{R}| = \kappa$. Note that if $i < \kappa$, then $\mathcal{R} \subseteq \mathcal{B}$ cannot be satisfied as $\mathcal{B} \overset{\$}{\leftarrow} \mathbb{B}_i$, and we have $H_\infty(\mathcal{B} \mid \text{false}) = \infty$ by default. For $i \geq \kappa$, it is an easy exercise to see that

$$H_\infty(\mathcal{B} \mid \mathcal{R} \subseteq \mathcal{B}) = -\log_2(\max_{b \in \mathbb{B}_i} \mathtt{Pr}[\mathcal{B} = b \mid \mathcal{R} \subseteq \mathcal{B}]) = -\log_2\left(\frac{1}{|\{\mathcal{B} \in \mathbb{B}_i \mid \mathcal{R} \subseteq \mathcal{B}\}|}\right)$$
$$= \log_2\left(|\{\mathcal{B} \in \mathbb{B}_i \mid \mathcal{R} \subseteq \mathcal{B}\}|\right) = \log_2\left(\binom{|\mathcal{U}| - \kappa}{i - \kappa}\right) =: \Phi_i\,,$$

as $\mathcal{R}$ fixes κ elements in $\mathcal{B}$, and the remaining $i - \kappa$ elements can be any of the $|\mathcal{U}| - \kappa$ remaining values in $\mathcal{U}$. □

The result reminds of the celebrated Erdös-Ko-Rado (EKR) theorem [10]. At a high-level, this theorem centers around the problem of determining maximum families $\mathbb{B}$ of λ-subsets of a set $\mathcal{U}$ such that any two sets have at least μ elements in common, and shows that $|\mathbb{B}| \leq \binom{|\mathcal{U}|-\mu}{\lambda-\mu}$ for large enough $\mathcal{U}$. Various variants of the EKR theorem have appeared, all differing in the conditions put on the set (see [8, 7] for a discussion). However, the seemingly apparent connection between the EKR theorem and Prop. 1 is merely coincidence, as in our case we focus on the (much simpler) problem of determining the number of sets $\mathcal{B}$ with $\mathcal{R} \subseteq \mathcal{B}$ for a given $\mathcal{R}$. Further, while the EKR theorem targets the derivation of an upper bound, we are effectively aiming for strong lower bounds on the size of $\mathbb{B}$.

4.2 Frientropy for Threshold Sets

A practically more interesting case is where $\mathbb{B} \subseteq \mathbb{P}(\mathcal{U}) \setminus \{\emptyset\}$ is incomplete, in which case the frientropy will naturally decrease. Indeed, if a user $u \in \mathcal{U}$ appears in exactly one set in $\mathcal{B} \in \mathbb{B}$, then $\mathbb{B}$ has 1-frientropy $\Phi_{|\mathcal{B}|} = 0$, which can be seen by taking $\mathcal{R} = \{u\}$. This leads to the definition of appearance in and the threshold of $\mathbb{B}$.

Definition 5. *Let $\mathbb{B} \subseteq \mathbb{P}(\mathcal{U}) \setminus \{\emptyset\}$. For a $u \in \mathcal{U}$, define its appearance in $\mathbb{B}$ by $\mathsf{appear}_{\mathbb{B}}(u) = |\{\mathcal{B} \in \mathbb{B} \mid u \in \mathcal{B}\}|$. Define the threshold of $\mathbb{B}$ as*

$$\tau_{\mathbb{B}} := \min_{\substack{u \in \mathcal{U}, \\ \mathsf{appear}_{\mathbb{B}}(u) > 0}} \mathsf{appear}_{\mathbb{B}}(u)\,. \tag{2}$$

Informally, the appearance of u in $\mathbb{B}$ is the number of sets in $\mathbb{B}$ that include u, and the threshold of $\mathbb{B}$ determines the minimal non-trivial appearance of any u.

We are now ready to derive the following result.

Proposition 2. *Consider any $\mathbb{B} \subseteq \mathbb{P}(\mathcal{U}) \setminus \{\emptyset\}$, and let $\mathcal{X}$ be the uniform distribution. Let $0 \leq \kappa \leq |\mathcal{U}|$. The set $\mathbb{B}$ has κ-frientropy $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_{|\mathcal{U}|})$, where*

$$\Phi_i = \begin{cases} \infty \text{ if } 1 \leq i < \kappa\,, \\ \log_2 \left(\max\{1, |\mathbb{B}_i| - (|\mathbb{B}_i| - \tau_{\mathbb{B}_i})\kappa\}\right) \text{ if } i \geq \kappa\,. \end{cases}$$

Proof. Let $\mathcal{R} \subseteq \mathcal{U}$ be a set of size $|\mathcal{R}| = \kappa$. The case $i < \kappa$ is as in Prop. 1. For $i \geq \kappa$, we can similarly compute

$$H_\infty(\mathcal{B} \mid \mathcal{R} \subseteq \mathcal{B}) = -\log_2(\max_{b \in \mathbb{B}_i} \mathtt{Pr}[\mathcal{B} = b \mid \mathcal{R} \subseteq \mathcal{B}]) = \log_2\left(|\{\mathcal{B} \in \mathbb{B}_i \mid \mathcal{R} \subseteq \mathcal{B}\}|\right)\,,$$

provided $|\{\mathcal{B} \in \mathbb{B}_i \mid \mathcal{R} \subseteq \mathcal{B}\}| > 0$. (Note that if this set were of size 0, we would necessarily have $\mathcal{R} \not\subseteq \mathcal{B}$ and thus $H_\infty(\mathcal{B} \mid \text{false}) = \infty$. We could henceforth discard this case.) This leaves us at the problem of determining a lower bound for the following problem: given κ elements $\mathcal{R}$, each of which appears in at least $\tau_{\mathbb{B}_i}$ sets $\mathcal{B} \in \mathbb{B}_i$, what is the number of sets $\mathcal{B} \in \mathbb{B}_i$ that include all elements from $\mathcal{R}$? A simple pigeonhole-principle-like computation shows that

$$|\{\mathcal{B} \in \mathbb{B}_i \mid \mathcal{R} \subseteq \mathcal{B}\}| \geq |\mathbb{B}_i| - (|\mathbb{B}_i| - \tau_{\mathbb{B}_i})\kappa\,,$$

as every of the κ elements in $\mathcal{R}$ eliminates at most $|\mathbb{B}_i| - \tau_{\mathbb{B}_i}$ sets in $\mathbb{B}_i$. Along with above-mentioned non-emptyness assumption on this set, we can put

$$\Phi_i := \log_2\left(\max\{1, |\mathbb{B}_i| - (|\mathbb{B}_i| - \tau_{\mathbb{B}_i})\kappa\}\right)$$

as stated. $\square$

A simple sanity check shows that if $\mathbb{B}$ is complete, the quantity in the $\log_2$-term of Prop. 2 reads $|\mathbb{B}_i| - (|\mathbb{B}_i| - \tau_{\mathbb{B}_i})\kappa = \binom{|\mathcal{U}|}{i} - \left(\binom{|\mathcal{U}|}{i} - \binom{|\mathcal{U}|-1}{i-1}\right)\kappa$, which is slightly worse than $\binom{|\mathcal{U}|-\kappa}{i-\kappa}$ of Prop. 1.

5 Frientropy in Practice

In [2] it is suggested that making the set $\mathcal{S}$ always of fixed size using dummy identities could provide some added security against privacy violations. This would mean that if the average size of the recipient set is some $n \in \mathbb{N}$, then during encryption the set $\mathcal{S}$ would be filled up to n with dummy identities or $\mathcal{S}$ would be divided into separate sets of size n and the last one would be filled with dummies. The suggestion is supported by Prop. 2, as it shows that the frientropy increases if for all i we have $\tau_{\mathbb{B}_i} \lesssim |\mathbb{B}_i|$. The adding of dummy friends could be done in such a way that the dummies are selected uniformly at random from the set $\mathcal{U}$ and added to the set $\mathcal{S}$. These dummy friends would not receive the "real" key, but a random string of bits.

However, this approach suffers from various caveats: first off, dummies would neither communicate nor participate in the discussion and the adversary could use this information to her advantage. Secondly, dummy identities should not and would not be able to access the contents, but this leads to the situation where those recipients selected as dummies would know that they are in fact dummies. In this case, the recipient privacy can only be assured for outsiders and not insiders. Care is needed to resolve these issues and support the use of dummy tweaks in a proper way. In addition, the task of the adversary would now be to distinguish real recipients from dummy ones, or if there are any dummy recipients in a given set of recipients.

Statistics show that an average OSN user on, e.g., Facebook, has 350 friends.[3] Furthermore, some statistics show that the average recipient set of group messages to be about 15 people. We remark, however, that the dynamic character of these networks causes the sets of recipients to vary over time, and this may influence the amount of frientropy and the level of privacy.

6 Membership Query Security

As noted earlier, the previous versions of recipient privacy do not take into account the nature of communications between possible recipients, and especially in the OSN setting this leads to diminishing privacy for the recipients. We will next present a stronger form of recipient privacy. It allows for the adversary to make also *membership* queries on the set of recipients in addition to the traditional decryption queries. A membership query tests whether a given u is in the set $\mathcal{S}$ used to encrypt the message.

One might think that this type of adversary is unrealistic and too powerful even in the OSN setting. However, as an example, think about a sysadmin on some content distribution system. In a well-maintained and -operated system, the sysadmin would probably have access to all kinds of logs and system information of even individual machines (of certain users). Thus, the sysadmin could a) know who has downloaded the content, and b) infer (through timing, power

[3] `http://www.statista.com/statistics/232499/americans-who-use-social-networking-sites-several-times-per-day` (read 2016-06-13)

consumption, or some other form of side-channel/social engineering) whether or not some of these people were able to correctly decrypt the content. In this sense, the power of such queries is not an unrealistic assumption for adversaries both in OSNs and in the original scenarios of recipient privacy.

Definition 6 (Membership Query Security). *Let $n \geq 1$, and let $\mathbb{B}_n \subseteq \mathbb{P}(\mathcal{U})$ be a set of subsets of $\mathcal{U}$ of size n. Let $\mathcal{X}$ be a probability distribution. The set provided membership query security if a PPT adversary $\mathcal{A}$ wins the following game with the challenger* `Ch`*, only with negligible probability:*

Init: `Ch` *chooses a set $\mathcal{S} \xleftarrow{\mathcal{X}} \mathbb{B}_n$.*

Phase 1: *$\mathcal{A}$ adaptively issues κ membership queries. A single membership query $M(u)$ reveals if $u \in \mathcal{S}$, i.e., $M(u) = \texttt{true}$ iff $u \in \mathcal{S}$ and* `false` *otherwise. Denote by $\mathcal{Q}$ the set of all users u for which $\mathcal{A}$ issues a query $M(u)$, and by $\mathcal{R}$ the set $\{u \in \mathcal{Q} : M(u) = \texttt{true}\}$.*

Guess: *$\mathcal{A}$ outputs a guess $\mathcal{S}' \in \mathbb{B}_n$.*

The adversary wins if $\mathcal{S}' = \mathcal{S}$, and we define $\mathsf{Adv}_{\mathcal{A}}^{\text{MemQ}} = \texttt{Pr}[\mathcal{S} = \mathcal{S}']$.

It is clear to see that membership query security is closely related to frientropy.

Proposition 3. *Let $\kappa, n \geq 1$. Consider any $\mathbb{B} \subseteq \mathbb{P}(\mathcal{U})$ consisting of subsets of $\mathcal{U}$ of size n. Assume that $\mathbb{B}$ has κ-frientropy $\boldsymbol{\Phi}$. Let $\mathcal{X}$ be the uniform distribution. For any adversary $\mathcal{A}$ making κ membership queries, we have $\mathsf{Adv}_{\mathcal{A}}^{\text{MemQ}} \leq 2^{-\Phi_\kappa}$.*

Proof. Note that any set of κ membership queries reveals a set of at most κ revealed users, as in the worst case all queries yield `true`. Denote these κ users by $\mathcal{R}$. The goal of the adversary $\mathcal{A}$ now is to guess $\mathcal{S} \in \mathbb{B}_n$ given that $\mathcal{R} \subseteq \mathcal{S}$. Its success probability is, by definition, at most $2^{-\Phi_\kappa}$. □

A more strict version would have the adversary guess for a single user $u \in \mathcal{S} \setminus \mathcal{R}$ and restrict that $\mathcal{R} \subset \mathcal{S}$ is a proper subset. It is easy to see that the notion of privacy against this adversary is stronger than that of the above game. The second way to amend the game would be to take into account the existence of dummy friends. The challenger would choose two sets $\mathcal{S}$ and $\mathcal{D}$ with $\mathcal{D} \subset \mathcal{S}$, where $\mathcal{D}$ represents the chosen dummies (if any). This would mean that the adversary would also output two sets $\mathcal{S}'$ and $\mathcal{D}'$ as guesses, with $|\mathcal{S}'| = |\mathcal{S}|$ and $\mathcal{D}' \subset \mathcal{S}'$, where $\mathcal{D}'$ is the adversary's guess for the dummies. In this game we can also model weak and strong membership queries, where the weak one is as in the above game and the strong gives the adversary also the information whether or not $u \in \mathcal{D}$, if $u \in \mathcal{S}$. The analysis of these variations is out of the scope of this paper and a possible venue for future research.

7 Conclusion

The traditional recipient privacy definitions fall short in providing provable security in OSN settings. The issue can be salvaged by requiring a high enough “frientropy,” meaning that if one of the recipients gets disclosed, the other recipients (its “friends”) remain mostly anonymous. Admittedly, sets of recipients are in practice not random; if a user u appears in only a few social groups, and it reveals itself as being a receiver of a message, the remaining set of recipients, conditioned on the fact that u appears in this group, has a very low entropy, and a way to resolve this is by using dummy friends.

References

1. Alwen, J., Dodis, Y., Wichs, D.: Leakage-resilient public-key cryptography in the bounded-retrieval model. In: CRYPTO 2009. LNCS, vol. 5677, pp. 36–54. Springer (2009)
2. Barth, A., Boneh, D., Waters, B.: Privacy in encrypted content distribution using private broadcast encryption. In: FC 2006. LNCS, vol. 4107, pp. 52–64. Springer (2006)
3. Beato, F.: Private Information Sharing in Online Communities. Ph.D. thesis, Katholieke Universiteit Leuven, Leuven (2015)
4. Beato, F., Ion, I., Capkun, S., Preneel, B., Langheinrich, M.: For some eyes only: protecting online information sharing. In: CODASPY’13. pp. 1–12. ACM (2013)
5. Beato, F., Kohlweiss, M., Wouters, K.: Scramble! your social network data. In: PETS 2011. LNCS, vol. 6794, pp. 211–225. Springer (2011)
6. Bellare, M., Boldyreva, A., Desai, A., Pointcheval, D.: Key-privacy in public-key encryption. In: ASIACRYPT 2001. LNCS, vol. 2248, pp. 566–582. Springer (2001)
7. Chung, F.R.K., Graham, R.L., Frankl, P., Shearer, J.B.: Some intersection theorems for ordered sets and graphs. Journal of Combinatorial Theory, Series A 43(1), 23–37 (1986)
8. Deza, M., Frankl, P.: Erdös-Ko-Rado theorem – 22 years later. SIAM Journal on Algebraic Discrete Methods 4(4), 419–431 (1983)
9. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In: PET 2002. LNCS, vol. 2482, pp. 54–68. Springer (2003)
10. Erdös, P., Ko, C., Rado, R.: Intersection theorems for systems of finite sets. The Quarterly Journal of Mathematics 12(1), 313–320 (1961)
11. Fazio, N., Perera, I.M.: Outsider-anonymous broadcast encryption with sublinear ciphertexts. In: PKC 2012. LNCS, vol. 7293, pp. 225–242. Springer (2012)
12. Günther, F., Manulis, M., Strufe, T.: Cryptographic treatment of private user profiles. In: RLCPS and WECSR 2011. LNCS, vol. 7126, pp. 40–54. Springer (2012)
13. Libert, B., Paterson, K.G., Quaglia, E.A.: Anonymous broadcast encryption: Adaptive security and efficient constructions in the standard model. In: PKC 2012. LNCS, vol. 7293, pp. 206–224. Springer (2012)
14. Pfitzmann, A., Köhntopp, M.: Anonymity, unobservability, and pseudonymity – A proposal for terminology. In: Designing Privacy Enhancing Technologies. LNCS, vol. 2009, pp. 1–9. Springer (2001)
15. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In: PET 2002. LNCS, vol. 2482, pp. 41–53. Springer (2003)
16. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557–570 (2002)
17. Tóth, G., Hornák, Z., Vajda, F.: Measuring anonymity revisited. In: NordSec 2004. pp. 85–90 (2004)