de-Linkability: a Privacy-Preserving Constraint for Safely Outsourcing Multimedia Documents

Bechara Al Bouna
Ticket Labs
Antonine University
Baabda, Lebanon
bechara.albouna@upa.edu.lb

Eliana J. Raad
LE2I-CNRS
Bourgogne University
Dijon, France
eliana.raad@u-bourgogne.fr

Charbel Elia
Ticket Labs
Antonine University
Baabda, Lebanon
charbel.elia@upa.edu.lb

Richard Chbeir
LIUPPA Laboratory
University of Pau and Adour Countries
Pau, France
rchbeir@acm.org

Ramzi Haraty
Department of Computer Science and Mathematics
Lebanese American University
Beirut, Lebanon
rharaty@lau.edu.lb

ABSTRACT

Outsourcing social multimedia documents is a growing practice among several companies in a way to shift their business globally. It is a cost-effective process where those companies tend to gain more profits disregarding eventual privacy risks. In fact, several case studies have showed that adversaries are capable of identifying individuals, whose identities need to be kept private, using the content of their multimedia documents. In this paper, we propose *de*-linkability, a privacy-preserving constraint to bound the amount of information outsourced that can be used to re-identify the individual. We also provide a sanitizing MD^*-algorithm to enforce *de*-linkability and present a set of experiments to demonstrate its efficiency.

Categories and Subject Descriptors

k.4.1 [**COMPUTERS AND SOCIETY**]: Public Policy Issues—*Privacy*

General Terms

Algorithms, Security

Keywords

Privacy, Multimedia Documents, de-Linkability

1. INTRODUCTION

Online social media and blogging are nowadays increasingly used to communicate with the wide range and diverse audience of the Web. In many situations, motivated by several campaigns such as politics, fraud fighting, cultural critics, and others, some authors of those social media need to remain anonymous. Consequently, when a data provider outsources multimedia documents, it becomes extremely hard sometimes to maintain individuals' anonymity mainly due, but not limited, to: 1) the number of active social networks to which they actually participate, and 2) the trails of seemingly information they leave behind [17]. These trails of information make individuals victims of what is known to the Internet community by *cyberstalking* where an adversary clandestinely tracks the movements of an individual. The "Twitter Hunt"[1] scenario in which an adversary was able to re-identify the previous french prime minister *François Fillon* expresses clearly the risk of re-identifying anonymous individuals. In this scenario, the adversary recognized the prime minister, who was using a fake account name "@fdbeauce" to remain anonymous, using previously published "tweets" which contained *enough* clues to disclose his identity. One of the clues that made this attack successful is the picture, and more precisely its metadata, of "*Château de Beaucé*" he published on his account uncovering the manor where he actually lives.

Indeed, exploiting inferable information can disclose anonymized identities where unrestricted access to online personal information remains a major threat. Most of the works done in the literature to preserve anonymity focus on structured relational data [26][23][8] while the only few techniques [10][15] proposed to handle identity anonymization in multimedia documents assume textual data with no reference whatsoever to multimedia objects such as images and videos.

In this paper, we propose *de*-linkability, a novel technique for preserving individual privacy when outsourcing multimedia documents. *de*-linkability ensures that individuals' identifiable information composed of both textual and multimedia content cannot be used to infer his/her identity.

Our contributions can be summarized as follows:

- We formally define the identity anonymization problem in multimedia documents composed of textual and multimedia

[1]http://www.euronews.com/2011/12/12/french-pms-shy-twitter-debut/

MEDES'13 October 28-31, 2013, Neumunster Abbey, Luxembourg

content. We use a selective intersection function to quantify the re-identification threat which is highly dependent on how much information can be acquired from 1) adversaries' background knowledge and 2) external sources containing relevant information related to the anonymized individual.

- We propose a *de*-linkability privacy constraint to bound individual re-identification due to textual and multimedia content that can be found in the outsourced multimedia documents.
- We present our sanitizing $\mathcal{MD}^*$-algorithm that allows to sanitize multimedia documents' content and preserve at the same time their utility in order to achieve the *de*-linkability.

The remainder of this paper is organized as follows. In Section 2, we present the adversary model adopted in our study. In Section 3, we discuss some of the works on anonymous document outsourcing and privacy preserving. Our data model definitions and operators are presented in Section 4. In Section 5, we give a formal definition of the re-identification problem. Section 6 is dedicated to present the *de*-linkability privacy constraint and to show how it is possible to preserve individual anonymity using a multimedia document sanitizing algorithm (the MD^*-algorithm). In Section 7, we evaluate our sanitizing algorithm to finally conclude and discuss our future research directions in Section 8.

2. ADVERSARY MODEL

In our adversary model, we assume that the adversary, that we call *cyberstalker*, knows that a given individual, that we call *cyberstalkee*[2], is hiding his/her identity (*François Fillon* in our scenario). We also assume that the *cyberstalker* has access to public information enabling him/her to link some personally identifying information, in a outsourced multimedia document, to the *cyberstalkee*. Thus, all *relevant* information (identifying or quasi-identifying) extracted from the document is considered individually identifiable.

More subtle, we assume that the *cyberstalker* has no prior knowledge of specific values for the stalked individuals. For example, the *cyberstalker* described in our motivating example does not know a-priori that "*Château de Beaucé*" is the residence of the *cyberstalkee*.

3. RELATED WORK

Several techniques have been defined in the literature [23] [26] [16] [14] to prevent information disclosure and eliminate possible linking attacks that are used for individual re-identification. These techniques assume that identifiable information and adversaries' background knowledge are stored in structured relational datasets. Specifically, they address linking attacks that can be established between (quasi)-identifying[3] and sensitive attributes of individuals stored in schema-based relational tables without referring to multimedia content.

Alternatively, techniques described in [10] and [15] preserve the individual's privacy in free text documents where data structure is missing. In [10], the authors measure sensitivity of identifiable information through a top-down propagation technique using prefixed sensitivity levels mapped to a reference ontology. According to these computed sensitivity levels, words are disseminated. In [15], the authors use a probabilistic-based algorithm to mine all searchable information concerning the individual. They use domain-specific ontologies to capture inferrable information and eventually provide more accurate results. Unfortunately, the ability of these techniques [10], [15] to deal with strong adversaries enforced with plausible background knowledge is limited when using domain-specific ontologies to compute sensitivity levels. These so-called levels of sensitivity should depend mainly on the knowledge that the adversary already has acquired which could be out of scope of a specific ontology. In [19], the authors propose a novel technique based on relevant occurrences to find user semantics. They assume that word co-occurrence is important to extract personal information from the Web. Similarly, in [11], the authors consider that the queries returning few results should be denoted as important. However, the amount of information is not always a relevant measure of dependency for privacy. For instance, two "tweets" with minimum co-occurrence might be issued by the same individual. Techniques described in [20], [24] and [5] are similar to a certain extent to our work. In [24], the authors propose a web-based solution to control undesired inferences. It first extracts relevant keywords from the document to be published and queries the web in order to capture additional knowledge contributing to a privacy breach. In [5], the authors present the notion of *k*-safety in which the identifying terms should be associated to at least *k* individuals. The authors in [20] sanitize sensitive parts of the document to measure information loss and risk disclosure. They assume that a relevant sanitizing process could be applied to maintain the utility of information in the document. As demonstrated in their experiments, these techniques are practical and promising, yet their ability to handle multimedia documents is limited. Unlike textual attributes, multimedia content cannot be approached without special processing to reduce uncertain decisions that overcome when similarity operators come to play. Here, we propose a technique to tackle individual re-identification threat caused by textual and multimedia content that can be linked to information obtained from external sources.

Table 1: Notations

u	an individual with anonymized identity
pf_u	an individual profile
mo	a multimedia object
$\mathcal{MD}_u$	a multimedia document related to u which should be sanitized
$\mathcal{MD}_\beta$	a multimedia document publicly accessible to adversaries extracted from an external source $\mathcal{E}$
$S_{\mathcal{MD}}$	a multimedia document signature
$\mathcal{E}$	an external source such as the social website, domain specific database, etc.
$\mathcal{C}$	a set of concepts
α	an association constant
β	an identification constant
ϖ	an aggregation function such as average, minimum, max, etc.

4. DATA MODEL

In this section, we define the data model and the basic notations (Table 1) used in the remainder of this paper.

4.1 Data Definition

Definition 1 (Multimedia Object)**.** *Let **mo** be any type of multimedia data such as an image, a video, or a salient object describing an object of interest (e.g., face of a person.). **mo** is formally repre-*

[2]Both terms *cyberstalkee* and *individual* will be used interchangeably in the remainder of this paper.

[3]http://www.springerreference.com/docs/html/chapterdbid/317693.html

sented as:

$$mo :\prec V, O, MO \succ \quad (1)$$

where:

- *V is a set of values describing the multimedia object. $\forall v_i \in V$ for $(1 \leq i \leq |V|)$ $v_i \in \mathcal{D}(e_i^t)$ where e_i^t is an attribute of type t and $\mathcal{D}(e_i^t)$ is the domain of e_i^t. We note that e_i^t can be any attribute of the Dublin Core Metadata Element set such that $t \in$ {source, description, date, contributor, format, etc.} or the MPEG-7 semantic set where $t \in$ {semantic place, concept, state, event, object}.*

- *O contains the raw data of the multimedia object. It can be a BFILE, an URL/URI, or an URL/URL augmented with a primitive to represent a salient object (e.g., Minimum Bounding Rectangle, Circle).*

- *MO is a set of multimedia objects directly contained in mo (For simplicity, we only consider multimedia objects at the first level of the hierarchy). We denote by $MO(mo_i)$ the set of multimedia objects contained in mo_i.*

For example, Figure 1 shows multimedia objects mo_{beauce} and mo_{manoir} representing two images of "*Château de Beaucé*" where *keywords* is an attribute of *mo*, *O* contains the raw data and MO is the empty set of multimedia objects contained in *mo*.

Definition 2 (Individual Profile)**.** *Given a cyberstalkee u, we denote by pf_u the profile of u formally defined as:*

$$pf_u :\prec PI, MO \succ \quad (2)$$

where:

- *PI is a set of values describing the individual's personal information. $\forall v_i \in PI$ for $(1 \leq i \leq |PI|)$, $v_i \in \mathcal{D}(a_i^t)$ where a_i^t is an attribute of type t and $\mathcal{D}(a_i^t)$ is the domain of a_i^t.*

- *MO is a set of multimedia objects attributed to u such that $\forall mo_i \in MO$ for $(1 \leq i \leq |MO|)$ $mo_i \in \mathcal{D}(ma_i^t)$. We note that ma_i^t is a multimedia attribute [4] of type t and $\mathcal{D}(ma_i^t)$ is the domain of ma_i^t.*

Referring back to our scenario, a typical profile of the previous french Prime Minister François Fillon would be:

$$pf_{Fillon} : (\text{"Francois Fillon","Prime Minister",} \\ \text{"}France\text{","}fcafillon@wanadoo.fr\text{"},mo_{beauce})$$

Definition 3 (Multimedia Document)**.** *Let $\mathcal{MD}$ be a multimedia document. $\mathcal{MD}$ is two dimensional and composed of a set of words and multimedia objects. It is formally defined as follows:*

$$\mathcal{MD} :\prec W, MO, \varsigma \succ \quad (3)$$

where:

- *W is a base text represented as a set of words where $m = |W|$.*

- *MO is set of multimedia objects contained in $\mathcal{MD}$ where $d = |MO|$.*

- *$\varsigma : MO \rightarrow W$ is a function that associates a multimedia object $mo_i \in MO$ for $(1 \leq i \leq d)$ to a word $w \in W$.*

[4] represents an attribute whose values are multimedia objects (e.g., pictureOf, imageOf, etc.)

An example of a multimedia document could be, but not limited to, personal blogs, set of tweets, newspaper articles, etc. Typically, these documents are composed of words and multimedia objects.

Now that we have defined our multimedia document, we present in the following what we call a multimedia document signature ($S_{\mathcal{MD}}$).

Definition 4 (Multimedia Document Signature)**.** *Let $\mathcal{MD}$ be a multimedia document, a multimedia document signature denoted by $S_{\mathcal{MD}}$ is a subset of $\mathcal{MD}$ composed of textual and multimedia content. $S_{\mathcal{MD}}$ is created using $S_{\mathcal{MD}} = IC(\mathcal{MD}, \mathcal{C})$ where IC is a function used to retrieve from $\mathcal{MD}$ relevant words and multimedia objects related to the set of concepts in $\mathcal{C}$.*

We assume that not all concepts found in a multimedia object provide meaningful clues that could lead to re-identify the *cyberstalkee*. For instance, it is unlikely for an individual working in a Health Care Department to be related to Computer Science. In other terms, some of the words and multimedia objects should more likely be related to the medical field instead of computing.

Using concepts to generate multimedia document signatures helps reducing the error rate of individual name disambiguation [12], particularly when the individual's profile is considered as a relevant source of concepts.

The followings are three sample multimedia documents' signatures generated based on the concepts *Country, Event* and *Location*.

$$S_{\mathcal{MD}_{Fillon}} : \\ (\text{"}Japan\text{","}Meeting\text{","@}beauce\text{"},mo_{beauce})$$

$S_{\mathcal{MD}_{Fillon}}$ is the anonymous multimedia document signature of Prime Minister Francois Fillon.

$$S_{\mathcal{MD}_{\beta_1}} : \\ (\text{"Francois Fillon","}France\text{","}Japan\text{","}Meeting\text{"})$$

$$S_{\mathcal{MD}_{\beta_2}} : \\ (\text{"Francois Fillon","}Home\text{"},mo_{manoir})$$

Both $S_{\mathcal{MD}_{\beta_1}}$ and $S_{\mathcal{MD}_{\beta_2}}$ are publicly available multimedia documents signatures related to Prime Minister Francois Fillon.

4.2 Data Manipulation

We provide in this section, the appropriate operators to address both multimedia and textual content of multimedia documents.

Definition 5 (Estimated Equality)**.** *Let W_1, W_2 be two sets of words over which an association function f can be used. Their estimated equality is computed as follows:*

$$equ(W_1, W_2) = \\ \varpi(f(w_1^1, w_1^2), \ldots, f(w_m^1, w_r^2)) \rightarrow [0,1] \quad (4)$$

where:

- *w_i^1, w_j^2 are two words of W_1 and W_2 respectively where $m = |W_1|$ and $r = |W_2|$.*

- *f is an association function defined as:*

$$f(w_i^1, w_j^2) = \begin{cases} 1 & \text{if } w_i^1 \in W_1 \text{ is the same as } w_j^2 \in W_2 \\ 0 & \text{otherwise} \end{cases}$$

mo_{beauce}	Keywords	O	MO
	Chateau de Beaucé	http://chateaubeauce.com/beauce.jpg	-

(a) Multimedia object representing "*Château de Beaucé*"

mo_{manoir}	Keywords	O	MO
	Manoir de Beaucé	http://manoibeauce.com/beauce.jpg	-

(b) Multimedia object representing "*Manoir de Beaucé*"

Figure 1: A typical description of two images using our multimedia object representation

- ϖ *is an aggregation function (e.g., max, min, avg, etc.) used to aggregate association functions' scores.*

The estimated equality is used to identify the amount of common textual values found in multimedia documents (or any subset of them). Alternatively, multimedia documents contain complex types such as images and videos which cannot be approached using traditional equality operators. We define in the following, an estimated similarity operator to process multimedia objects.

Definition 6 (Estimated Similarity)**.** *Let MO_1, MO_2 be two sets of multimedia objects over which n similarity functions $s_1,\ldots,s_n$ can be used. Their similarity score is computed as follows:*

$$sim(MO_1,MO_2) = \varpi(s_1(mo_1^1,mo_1^2),\ldots,s_n(mo_m^1,mo_r^2)) \rightarrow [0,1] \quad (5)$$

where:

- mo_i^1, mo_j^2 *are two multimedia objects of MO_1 and MO_2 respectively where $m = |MO_1|$ and $r = |MO_2|$.*
- s_k *is a unit similarity function comparing multimedia objects $mo_i^1 \in MO_1$ and $mo_j^2 \in MO_2$. We note that $s_k(mo_i^1,mo_j^2)$ compares mo_i^1, and mo_j^2 based on their attributes and raw data* [5]*. s_k returns a score between [0, 1], where 0 expresses a total divergence and 1 a complete similarity.*
- ϖ *is an aggregation function used to aggregate the computed similarity scores.*

We show in the following how multimedia documents intersection can be determined using selective intersection.

Definition 7 (Selective Intersection)**.** *Let $S_{\mathcal{MD}_1}$, $S_{\mathcal{MD}_2}$ be two distinct multimedia documents signatures, their selective intersection is defined as:*

$$S_{el}I_{nt}(S_{\mathcal{MD}_1},S_{\mathcal{MD}_2}) = ||\sum_{c_i} a_{c_i} \times equ_{c_i}(W_1,W_2) + \sum_{c_j} a_{c_j} \times sim_{c_j}(MO_1,MO_2)|| \quad (6)$$

where:

- *c represents a concept for which an equality and/or similarity should be computed. Such concepts, either user defined or retrieved based on their relevance in the multimedia document, can be used to selectively choose relevant content in multimedia documents. For instance, it is possible to capture the amount of common information related to the concept Person. This refers to computing the equality and similarity of words and multimedia objects that are related to the concept Person for both multimedia document signatures $S_{\mathcal{MD}_1}, S_{\mathcal{MD}_2}$.*
- a_c *is the priority assigned to each concept c where its magnitude depends on the normalizing assumptions.*

Selective intersection returns a normalized score $\in [0,1]$ computed based on equality and similarity of multimedia documents content. For instance, let us compute the selective intersection between $S_{\mathcal{MD}_{Fillon}}$ and both $S_{\mathcal{MD}_{\beta_1}}$ and $S_{\mathcal{MD}_{\beta_2}}$. We adopt the *max* aggregation function to compute the equality and/or similarity scores for each concept and finally determine their average score. The selective intersection based on concepts *Country, Event* and *Location* is detailed below:

$$S_{el}I_{nt}(S_{\mathcal{MD}_{Fillon}},S_{\mathcal{MD}_{\beta_1}}) = \frac{\frac{1+1+0}{3}+0}{2} = 0.33$$
$$S_{el}I_{nt}(S_{\mathcal{MD}_{Fillon}},S_{\mathcal{MD}_{\beta_2}}) = \frac{\frac{0}{3}+0.8}{2} = 0.4$$

We assume, in this case, that the estimated similarity between multimedia objects mo_{beauce} and mo_{manoir} in $S_{\mathcal{MD}_{Fillon}}$ and $S_{\mathcal{MD}_{\beta_2}}$ returned a 0.8 score.

Unlike mutual information metric [22], our selective intersection is a non-correlation based metric where the count of each value in the signatures has minimum influence on the overall computation

[5] we invite the reader to consult our work on multimedia objects similarity computation in [2]

score. Specifically and for privacy reasons, this assumption is useful to determine minimum intersection between multimedia documents. We will show in the following definition, the premise of multimedia documents association.

Definition 8 (α-*association*)**.** *Let $\mathcal{MD}_1, \mathcal{MD}_2$ be two distinct multimedia documents. We say that an α-association exists between $\mathcal{MD}_1$ and $\mathcal{MD}_2$ if their selective intersection $SelInt(S_{\mathcal{MD}_1}, S_{\mathcal{MD}_2})$ is greater than $1/\alpha$ where:*

- *$S_{\mathcal{MD}_1}$ and $S_{\mathcal{MD}_2}$ represent corresponding multimedia documents signatures.*
- *$\alpha \geq 2$ is the association constant.*

α-*association* expresses the presence of a possible association between two multimedia documents represented by their signatures. It measures the strength of an association between two multimedia document signatures based on mutual information composed of both textual and multimedia content.

5. IDENTITY ANONYMIZATION PROBLEM

In the presence of adversaries with sophisticated tracking abilities, privacy and ownership preserving of outsourced data tends to be a complex task. Such adversaries, armed with plausible background knowledge and a wide range of accessible web-based social information, compromise anonymization techniques and put at risk individuals' privacy. Here, we express the identity anonymization problem that could arise when outsourcing multimedia documents as the amount of information accessible by the adversary and that can be, at the same time, associated to the owner of the outsourced multimedia documents. It is formally defined as follows:

Definition 9 (Identity Anonymization Problem)**.** *Let $\mathcal{MD}_u$ be the multimedia document of an individual u. We say that an adversary is able to re-identify u from $\mathcal{MD}_u$ if $\exists \mathcal{MD}_\beta$, a publicly available multimedia document, such that:*

1. *$\mathcal{MD}_u$ and $\mathcal{MD}_\beta$ are α-associated and,*
2. *The knowledge related to u that can be obtained from $\mathcal{MD}_\beta$ is greater than $1/\beta$. It is expressed as a β-association between $\mathcal{MD}_\beta$ and the individual profile pf_u where α is the association constant, $\beta \geq 2$ is an identification constant and both are user-defined.*

It is difficult to know how much the adversaries know and to what extent their ability to disclose individuals' identities can be compromising. Here, we only avoid leaking information to the *cyberstalker* except for what he/she already has. Such assumption is no different than the one adopted by differential privacy [8] where our main objective is essentially providing constraints on the release of the data.

6. PRIVACY PRESERVING

Preserving privacy requires that the *cyberstalker* remains incapable of identifying the anonymized identity of the *cyberstalkee*, owner of the multimedia document to be published. As we have stated in the previous section, a re-identification threat occurs mainly due to:

- the link between his/her related multimedia document $\mathcal{MD}_u$ and a multimedia document $\mathcal{MD}_\beta$ accessible by the *cyberstalker* and,
- the amount of information extracted from $\mathcal{MD}_\beta$ and associated to *u*.

Controlling the latter can be a burden or eventually unrealizable due to accessibility issues while, on the other hand, breaking the link between multimedia documents is achievable and can be done using *de*-linkability.

***de*-linkability.** *Given a cyberstalkee u and a multimedia document $\mathcal{MD}_u$, the de-linkability privacy-preserving constraint is satisfied if $\forall \mathcal{MD}_\beta \in \sigma_{E_u}(\mathcal{E})$ that is β-associated to pf_u, $\mathcal{MD}_u$ cannot be linked to $\mathcal{MD}_\beta$ through an α-association, where $\sigma_{E_u}(\mathcal{E})$ is a selection on an external source $\mathcal{E}$ based on a conjunctive set of words and/or multimedia objects (E_u) related to u.*

de-linkability breaks the link between an outsourced multimedia document and any other document *accessible* to a *cyberstalker* and that can be linked to *u*. It is important to note that the content of E_u that is used to retrieve multimedia documents $\mathcal{MD}_\beta$ from the external source should be considered carefully in order to reduce the scope of potential error. A straightforward assumption is to consider this content as a subset of the individual's profile including both identifying and quasi-identifying values.

6.1 Achieving *de*-linkability

In keeping with many works in anonymization, *de*-linkability can be achieved using a straightforward extension of traditional anonymization techniques such as suppression, substitution or generalization relationships between domains and values [23][26][25] for textual values in $\mathcal{MD}_u$ as long as there is no $\mathcal{MD}_\beta$ that can be α-associated to $\mathcal{MD}_u$. Unsurprisingly, multimedia objects need a special interest. Eventually, the objective is to break linkable objects that could contribute in re-identifying the anonymized individual. More subtle is to hide and/or disseminate multimedia objects content while at the same time preserving a minimum semantic or visual coherence.

In this paper, we do not provide a in-depth details on how multimedia objects content could be protected. This matter is left for future work. We only use traditional techniques to protect salient objects as in [3] where the authors protect textual and image data through flexible low-level adapted security rules, while in [9] object substitution is adopted. In [4], blurring proved efficiency, and objects removal from images and videos were addressed in [7] [13] [27] [28] [6] [21].

Here, we refer to this process as document sanitizing which we formally define as follows:

Definition 10 (Multimedia Document Sanitizing)**.** *Let $\mathcal{MD}_u$ be the multimedia document related to a cyberstalkee u. Given $\widetilde{\mathcal{G}}_W$ and $\widetilde{\mathcal{G}}_{MO}$ two corresponding sanitizing functions, we say that $\mathcal{MD}_u$ is sanitized, denoted by $\mathcal{MD}^*_u = \widetilde{\mathcal{G}}_{(W,MO)}(\mathcal{MD}_u)$ if both words and multimedia objects are sanitized $\widetilde{\mathcal{G}}_W(W_{MD_u})$ and $\widetilde{\mathcal{G}}_{MO}(MO_{MD_u})$.*

Multimedia document sanitizing ensures that the specified content (W, MO) is either removed, suppressed, generalized and/or protected in the multimedia document $\mathcal{MD}_u$ based on the sanitizing function $\widetilde{\mathcal{G}}$.

6.2 Multimedia Document Sanitizing: $\mathcal{MD}^*$ – *algorithm*

$\mathcal{MD}^*$-algorithm is used to sanitize a multimedia document and protect the *cyberstalkee*'s identity. As mentioned in the pseudocode, the algorithm takes a multimedia document $\mathcal{MD}_u$, a set of

Algorithm 1 $\mathcal{MD}^*$-algorithm

Require: a multimedia document $\mathcal{MD}_u$, set of concepts $\mathcal{C}$ over $\mathcal{MD}_u$, an individual profile pf_u, conjunctive set of words and/or multimedia objects E_u, association constant α and identification constant β
Ensure: Multimedia Document Sanitizing $\mathcal{MD}^*_u$

1: $S_{\mathcal{MD}_u} = IC(\mathcal{MD}_u, \mathcal{C})$ ▷ Generate Multimedia Signature on $\mathcal{MD}_u$
2: **for each** $\mathcal{MD}_\beta$ **in** $\sigma_{E_u}(\mathcal{E})$ **do**
3: $S_{\mathcal{MD}_\beta} = IC(\mathcal{MD}_\beta, \mathcal{C})$ ▷ Generate Multimedia Signature on $\mathcal{MD}_\beta$
4: **if** $S_{el}I_{nt}(S_{\mathcal{MD}_\beta}, pf_u) > \frac{1}{\beta}$ **then**
5: **while** $S_{el}I_{nt}(S_{\mathcal{MD}_u}, S_{\mathcal{MD}_\beta}) > \frac{1}{\alpha}$ **do**
6: Retrieve least significantly threatening W_β and MO_β
7: $\mathcal{MD}^*_u \leftarrow \widetilde{\mathcal{G}}_{(W_\beta, \vec{MO}_\beta)}(\mathcal{MD}_u)$ ▷ Sanitize $\mathcal{MD}_u$ based on W_β and MO_β
8: **end while**
9: **end if**
10: **end for**

concepts $\mathcal{C}$ related to u (used to extract multimedia document signature), the *cyberstalkee* profile pf_u along with E_u and both association and identification constants α, β. It returns a sanitized multimedia document ($\mathcal{MD}^*_u$).

The $\mathcal{MD}^*$-algorithm extracts in Step 1 the multimedia document signature $S_{\mathcal{MD}_u}$ using the extraction function IC. It sanitizes $\mathcal{MD}_u$ from Step 2 to 10.

In Step 3, it extracts the signature of a multimedia document $\mathcal{MD}_\beta$ retrieved from an external source $\mathcal{E}$ based on the set of entities E_u related to u. In order to determine the amount of information related to u and that can be obtained from $\mathcal{MD}_\beta$, we compute the selective intersection on $\mathcal{MD}_\beta$ and the *cyberstalkee* profile pf_u. If their selective intersection $S_{el}I_{nt}(S_{\mathcal{MD}_\beta}, pf_u)$ is greater than $1/\beta$, the link between $\mathcal{MD}_u$ and $\mathcal{MD}_\beta$ should be anonymized as done from Step 5 to 8. That is, as long as they are α-associated the **least**[6] significant W_β and MO_β are sanitized in $\mathcal{MD}_u$.

The $\mathcal{MD}^*$-algorithm's time complexity can be estimated to $O(|\mathcal{MD}_\beta| \times (|W^*| + |MO^*|))$ where $|\mathcal{MD}_\beta|$ is the number of relevant multimedia documents retrieved from the external source and $|W^*| + |MO^*|$ is the number of sanitized words and multimedia objects from MD_u.

7. EXPERIMENTS

In this section, we present a set of experiments to evaluate the efficiency of our approach. We implemented the $\mathcal{MD}^*$-algorithm code[7] in Java and conducted experiments using a 3.4 GHz Intel Core i7 with 16 GB RAM.

7.1 Dataset Configuration

We used 200 individuals of the dataset published[8] by the authors of [1]. For each individual, we grouped 100 of his/her tweets to form his/her $\mathcal{MD}_u$. These $\mathcal{MD}_u$ have been filtered to remove identifying names. OpenCalais api[9] is used to extract "*concepts*" from multimedia documents $\mathcal{MD}_u$ and $\mathcal{MD}_\beta$. We actually used the most relevant concepts extracted based on a predefined threshold that we have set to 0.5 (this threshold can be used to fine-tune the evaluation results and include relevant concepts).

Alternatively, we used images to refer to multimedia objects in particular, we adopted the Zemanta api[10] to retrieve and associate images with their related words contained in $\mathcal{MD}_u$. This enables us to determine the content in $\mathcal{MD}_u$ that could be linked to *external images* that represent a possible re-identification threat.

Individual profiles pf_u were downloaded using the Twitter api[11]. For our assessment, we only focused on four profile attributes namely *name, screen name, location and profile_image_url*.

As per *cyberstalkee*, we retrieved up to 10 relevant multimedia documents $\mathcal{MD}_\beta$ using the Google api[12] applying to the individual *name* combined to relevant content from his/her related $\mathcal{MD}_u$. This way, we can assert that the retrieved multimedia documents $\mathcal{MD}_\beta$ are related to the *cyberstalkee* at hand at least through their names.

To compare images, we used the phash function[13] and assigned a weight of 0.5 to the estimated similarity for the selective intersection $S_{el}I_{nt}$.

7.2 Evaluation Results

We elaborated a set of measurements to evaluate the efficiency of the $\mathcal{MD}^*$-algorithm. These measurements can be summarized as follows:

- Evaluating the identity anonymization problem represented by the percentage of individuals re-identified.
- Determining sanitizing cost to capture the content that should be sanitized in order to prevent a privacy violation.
- Determining the computational cost of our $\mathcal{MD}^*$-algorithm.

7.2.1 *Evaluating Privacy*

In this test, we evaluated the identity anonymization problem represented by the percentage of individuals identified according to what they have published in their $\mathcal{MD}_u$ and their related multimedia documents $\mathcal{MD}_\beta$. We fixed the identification threshold $\beta = 10$ in order to capture significant number of multimedia documents related to individual u and used various association thresholds $\alpha = 2, 4, 6, 8$ and 10. The results shown in Figure 2 show the percentage of re-identified individuals from the total number of individuals processed.

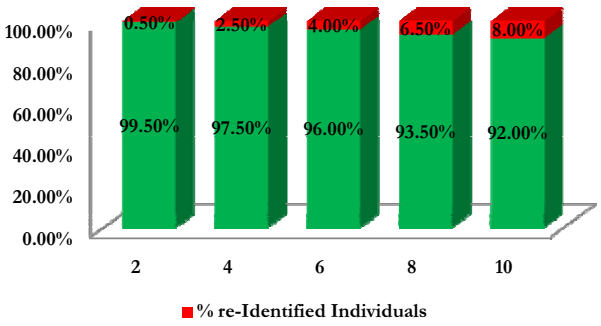

Figure 2: Privacy Violation Evaluation

[6]The importance of retrieved W_β and MO_β is determined based on the priority thresholds prefixed in the selective intersection function.
[7]The source code of the prototype can be downloaded from http://sourceforge.net/p/pmi1/code/HEAD/tree/trunk/MDanon/
[8]http://wis.ewi.tudelft.nl/umap2011/
[9]http://www.opencalais.com/
[10]http://developer.zemanta.com/
[11]https://dev.twitter.com/
[12]https://developers.google.com/custom-search/v1/overview
[13]http://phash.org/docs/howto.html

We can see that when the association threshold increases, there is a higher chance of linking individuals to the multimedia documents $\mathcal{MD}_\beta$ retrieved from the external source and eventually leading to their re-identification.

7.2.2 *Evaluating Sanitizing Cost*

We evaluate the MD^*-algorithm to determine the anonymization cost and estimate the increasing uncertainty due to the sanitizing process[14]. To do so, we calculate the average entropy [18] of individuals' multimedia documents $\mathcal{MD}_u$ in a pre- and post-sanitizing process. As a matter of fact, for each individual's multimedia document, we compute its entropy based on the concepts used to generate its own multimedia document signature (see Definition 4) as:

$$Entropy(\mathcal{MD}_u) = -\sum_{c \in C} Pr(c) log(Pr(c))$$

where c is the related concept.

We estimate the uncertainty to be: $|Entropy(\mathcal{MD}_u) - Entropy(\mathcal{MD}_u^*)|$ where $\mathcal{MD}_u^*$ is the sanitized individual's multimedia document. The results are shown in Figure 3.

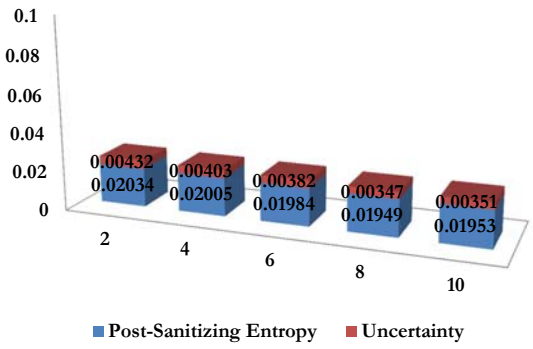

Figure 3: Sanitizing Cost Evaluation

Figure 3 shows that the uncertainty caused by the sanitizing process is relatively small. This uncertainty could get even smaller if sanitizing multimedia objects was approached differently using blurring or pixelizing techniques that preserve the semantic and coherence of images' content. This process is left for a future work.

7.2.3 *Evaluating Computational Cost*

We evaluated here the computational cost of the $\mathcal{MD}^*$-algorithm. We can see in Figure 3 that the sanitizing process can be achieved in a polynomial time.

[14]we have omitted the threatening values and objects from our evaluation process

Figure 4: Computational Cost Evaluation

The resulting computational time depends on: 1) the conjunctive set of words and/or multimedia objects in E_u that are used to query the external source, 2) the external source from which multimedia documents ($\mathcal{MD}_\beta$) are retrieved (e.g., the Web in our case). This is what we call fetching time which in some cases can be unpredictable as noticed between $\alpha = 4$ to 6 where the time to retrieve the individuals' data from the external source has increased.

8. CONCLUSION

Multimedia documents outsourcing has become part of the routine activity of many social companies. Such data sharing puts at risk the privacy of individuals when adversaries get the ability to associate the multimedia document's content to possible trail of information left behind by the individual. In this paper, we showed how this breach can be achieved and proposed *de*-linkability to cope with it. *de*-linkability is a privacy-preserving constraint that ensures the safe outsourcing of multimedia documents to semi-trusted third parties. It deals with the privacy threat in its broader aspect while considering both textual and multimedia content. We provided a sanitizing algorithm to protect against violating content and preserve at the same time a minimum quality through an adapted sanitizing process that takes into consideration the complex nature of multimedia objects. In the near future, we expect to provide more tests to demonstrate the efficiency of *de*-linkability. We also intend to extend our technique to include an in-depth quality assessment and evaluation for both multimedia and textual attributes.

9. ACKNOWLEDGEMENT

This study is funded by the Lebanese CNRS Research Grant Program NCSR project 506 fund 1003. It is also partly funded by the CEDRE research collaboration program, project AO 2011, entitled: Easy Search and Partitioning of Visual Multimedia Data Repositories, jointly funded by the French CNRS (National Center for Scientific Research)

10. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendations. In J. Konstan, R. Conejo, J. Marzo, and N. Oliver, editors, *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2011.

[2] B. Al Bouna, R. Chbeir, and S. Marrara. A multimedia access control language for virtual and ambient intelligence environments. In *Proceedings of the 2007 ACM workshop on Secure web services*, SWS '07, pages 111–120, New York, NY, USA, 2007. ACM.

[3] B. A. Bouna, R. Chbeir, and A. Gabillon. The image protector - a flexible security rule specification toolkit. In *SECRYPT*, pages 345–350, 2011.

[4] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *CSCW*, pages 1–10, Philadelphia, Pennsylvania, 2000. ACM.

[5] V. T. Chakaravarthy, H. Gupta, P. Roy, and M. K. Mohania. Efficient techniques for document sanitization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 843–852, New York, NY, USA, 2008. ACM.

[6] B. T. Chun, Y. Bae, and T.-Y. Kim. A method for original image recovery for caption areas in video. In *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, volume 2, pages 930 – 935, 1999.

[7] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *Image Processing, IEEE Transactions on*, 13(9):1200 –1212, sept. 2004.

[8] C. Dwork. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.

[9] J. Fan, H. Luo, M.-S. Hacid, and E. Bertino. A novel approach for privacy-preserving video sharing. In *CIKM*, pages 609–616, Bremen, Germany, 2005. ACM.

[10] L. Geng, Y. You, Y. Wang, and H. Liu. Privacy measures for free text documents: Bridging the gap between theory and practice. In *TrustBus*, pages 161–173, 2011.

[11] E. Gessiou, Q. H. Vu, and S. Ioannidis. Irild: An information retrieval based method for information leak detection. In *Computer Network Defense (EC2ND), 2011 Seventh European Conference on*, pages 33–40, 2011.

[12] J. Huang, S. Ertekin, and C. L. Giles. Efficient name disambiguation for large-scale databases. In *PKDD*, pages 536–544. Springer-Verlag, 2006.

[13] N. Komodakis. Image completion using global optimization. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 442 – 452, june 2006.

[14] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE*, pages 106–115, 2007.

[15] R. Ma, X. Meng, and Z. Wang. Preserving privacy on the searchable internet. In *iiWAS*, pages 238–245, 2011.

[16] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. L-diversity: privacy beyond k-anonymity. In *Data Engineering, 2006. ICDE '06. Proceedings of the 22nd International Conference on*, pages 24–24, 2006.

[17] B. Malin. *Trail Re-identification and Unlinkability in Distributed Databases*. PhD thesis, Carnegie Mellon University, 2006.

[18] N. F. Martin and J. W. England. *Mathematical theory of entropy*, volume 12. Cambridge University Press, 2011.

[19] J. Mori, Y. Matsuo, and M. Ishizuka. Finding user semantics on the web using word co-occurrence information. In *Proceedings of the International Workshop on Personalization on the Semantic Web (PersWeb05)*, 2005.

[20] D. F. Nettleton and D. Abril. Document sanitization: Measuring search engine information loss and risk of disclosure for the wikileaks cables. In J. Domingo-Ferrer and I. Tinnirello, editors, *Privacy in Statistical Databases*, volume 7556 of *Lecture Notes in Computer Science*, pages 308–321. Springer Berlin Heidelberg, 2012.

[21] K. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting under constrained camera motion. *Image Processing, IEEE Transactions on*, 16(2):545 –553, feb. 2007.

[22] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007.

[23] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. Knowl. Data Eng.*, 13(6):1010–1027, 2001.

[24] J. Staddon, P. Golle, and B. Zimny. Web-based inference detection. In *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*, SS'07, pages 6:1–6:16, Berkeley, CA, USA, 2007. USENIX Association.

[25] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):571–588, 2002.

[26] L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[27] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, june 2008.

[28] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(3):463 –476, march 2007.