IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 12, 2019, accepted July 25, 2019, date of publication August 2, 2019, date of current version August 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932899

IPAM: Information Privacy Assessment Metric in Microblogging Online Social Networks

SAMIA OUKEMENI[1], HELENA RIFÀ-POUS[1], AND JOAN MANUEL MARQUÈS PUIG[2]

[1] Internet Interdisciplinary Institute (IN3), CYBERCAT-Center for Cybersecurity Research of Catalonia, Universitat Oberta de Catalunya (UOC), 08018 Barcelona, Spain

[2] Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya (UOC), 08018 Barcelona, Spain

Corresponding author: Samia Oukemeni (soukemeni@uoc.edu)

This work was supported in part by the Spanish Government (CONSENT) under Grant RTI2018-095094-B-C22 and (SMARTGLACIS) under Grant TIN2014-57364-C2-2-R.

ABSTRACT A large amount of sensitive data is transferred, stored, processed, and analyzed daily in Online Social Networks (OSNs). Thus, an effective and efficient evaluation of the privacy level provided in such services is necessary to meet user expectations and comply with the requirement of the applicable laws and regulations. Several prior works have proposed mechanisms for evaluating and calculating privacy scores in OSNs. However, current models are system-specific and assess privacy only from the user's perspective. There is still a lack of a universal model that can quantify the level of privacy and compare between different systems. In this paper, we propose a generic framework to (i) guide the development of privacy metrics and (ii) to measure and assess the privacy level of OSNs, more specifically microblogging systems. The present study develops an algorithmic model to compute privacy scores based on the impact of privacy and security requirements, accessibility, and difficulty of information extraction. The proposed framework aims to provide users as well as system providers with a measure of how much the investigated system is protecting privacy. It allows also comparing the privacy protection level between different systems. The privacy score framework has been tested using real microblogging social networks and the results show the potential of the proposed privacy scoring framework.

INDEX TERMS Online social networks, information privacy protection, information security, metrics, microbloggings, privacy measurements.

I. INTRODUCTION

online Social Networks (OSNs) offer a free channel for users to express themselves, to share their thoughts and activities, and to communicate with others. OSNs are not limited to a specific demographic or age group, but they encourage people of all ages and demographics worldwide to participate by sharing their thoughts and interests or by advertising their products. Currently, it is estimated that around 2.77 billion users are actively using OSNs platforms [1]. Furthermore, active participation in OSNs is shaping the way people communicate. It is quite remarkable how news spread so fast on social networks. For example, OSNs have played a crucial role in the recent events of the 'Arab Spring' [2] or the 'London riots' [3].

On the other side, OSNs users provide a variety of personal information that may disclose more sensitive information about the users or their entourage and cause potential privacy risks. For example, the website Please Rob Me [4] was created to raise the awareness of the dangers of location-based services on Twitter. It scans tweets and shows when the users tweet from places other than their home.

As a solution, OSNs provide privacy policies and privacy settings to control and adjust who can access users' profiles and posts [5], [6]. However, privacy policies offered by the systems are confusing and expressed in legal jargon that is difficult to understand [7]. Furthermore, privacy settings are complex, time-consuming, and still insufficient to fully protect users' privacy [8]. Besides, OSNs providers mostly store, process and analyze users' data and sometimes sell them to third-parties for advertising and marketing purposes [9], [10].

As the monetary value of data increases, more voices demand protection and control over their privacy. The year 2018 saw a trend to enact new laws that regulate data collection and enhance privacy protection. For example, the European Parliament approved the General Data Protection

The associate editor coordinating the review of this article and approving it for publication was Chunsheng Zhu.

Regulation (GDPR) on 14 April 2016 and to be applied by 25 May 2018 [11]. The new regulation aims to give control to online users over their personal data, to harmonize data privacy laws across Europe, and to protect the privacy of the users [12]. However, GDPR is not the miraculous solution to protect the users' privacy. The end-users have to accept the terms and conditions as provided by the system, they are not in a position to negotiate a separate agreement. Therefore, the users are obliged to accept to release their data as a necessity to use the services [13]. Moreover, the regulation was criticized for the lingering uncertainty around some undefined terms (e.g. ''disproportionate effort'' or ''undue delay'') which require more clarity by the courts and regulators. Furthermore, GDPR does not offer a proper definition of what constitutes a ''reasonable'' level of protection for personal data, offering flexibility in the assessment of fines for data breaches and non-compliance [14].

The public interest in privacy protection has increased due to the growing amount of data breaches of common and extended use services. Privacy-oriented service providers and researchers have introduced new systems that offer online social functionalities and at the same time advocate for privacy protection. Some systems add a privacy layer while others are built using privacy by design methodologies.

With a multitude of privacy controls and techniques implemented in these new microblogging OSNs, the necessity of tools to evaluate the privacy protection techniques in OSNs has appeared. Therefore, there is a need for a formal framework that quantifies privacy and evaluates the performance and the efficiency of the privacy-preserving techniques implemented in OSNs. The scores obtained from using the framework can provide users with a means to compare different OSNs and to choose the most adequate for them depending on their risk appetite. In this paper, the focus is on OSNs that offer microblogging service. Privacy evaluation can establish credibility and trust in social network services and incite platforms' providers to address privacy concerns.

The remaining proceeds as follows: section II presents the current state of the art of OSNs and privacy; section III discusses the existing privacy-specific measurement models in OSNs, while section IV presents the gap between the surveyed models, and explains the contributions of this paper; section V presents the methodology of the proposed framework and section VI presents the algorithmic model for calculating privacy score in microblogging OSNs, and an evaluation and assessment of the framework based on real social networks are discussed in section VII, while Section VIII presents a comparison between our framework and related work. Section IX presents the conclusion of the paper, including future work.

II. PRIVACY AND OSNs

A. ONLINE SOCIAL NETWORKS

An Online Social Network (OSN) is a translation of physical connections and relationships to the virtual world. In general, it is defined as user-generated content services that facilitate social interactions through the internet. OSN allows (i) to create public or semi-public profiles, (ii) to build social relationships with other users, and (iii) to post their activities and interests and view those made by others [15]. Microblogging network (MOSN) is a popular form of OSNs. It is a weblog where users can send snippets of a small number of characters (between 140 and 310 characters) [16].

An OSN can be modeled as a graph where nodes are users and edges are the relationships between the users. The edges can be either unidirectional (e.g. Twitter social model of following) or bidirectional (e.g. Facebook social model of friendship) [17], [18].

When studying MOSNs, there are two major aspects that characterize them: the targeted stakeholders involved in the usage of the microblogging services and the data collected and used in the system [19], [20].

1) MOSNs STAKEHOLDERS

MOSNs stakeholders are defined as entities that can access user-data directly or indirectly. They can be categorized into users, service operators, third-parties, and the general public.

- **User**: is any entity (an individual or an organization) that subscribes to a MOSN to benefit from the services offered [19]. A user is represented by a profile, her relationships, and her generated contents.
- **System operator** (also known as the service provider): provides the underlying services and infrastructure needed to use the system and interact with each other [19]. They play the role of data protectors, and they can have direct or indirect access to the user data.
- **Third-parties**: are entities that connect to either system operators or users for different objectives other than social networking [19]. They can be developers of applications, providers of add-ons, data analysts, marketers, advertisement agencies, etc.
- **General public**: are unregistered users that can access (if they are allowed) to the services of the OSNs to visualize, monitor, or extract information.

2) DATA IN MOSNs

When using microblogging services, users release personal information either willingly or unintentionally. User data are either provided in the user's profile, generated by the user, shared in groups, collected from patterns, or derived from all other types of data. Kumari, in [19], classifies user data into two:

- **Traffic data**: All data generated from users' activities, such as IP address, OS specifications, search terms, etc.
- **Payload data**: All data posted by the user herself or about the user and by other users.

Furthermore, the payload data can be classified into [21]:

- **Service data**: are the data the users provide to use the services of the social networking site, like name, age, email address…

- **Disclosed data**: are the data generated by the users in their pages.
- **Entrusted data**: are the data that a user posts on other people's pages, like commenting on friends' messages, photos, and videos, tagging friends in photos, etc.
- **Incidental data**: are the data posted by other users about the user. It is similar to the entrusted data, but the user has no control over the data posted. They can include comments on the user's photos and videos, comments on the user's status updates', etc.
- **Behavioral data**: are the data collected about the users' habits and activities.
- **Hidden or derived data**: are the data derived from all the other types of data.

B. PRIVACY DEFINITION

The concept of privacy is complex and there is no mere universal definition. It can be interpreted in different ways and contexts, and it varies depending on the discipline perspectives, i.e. law, health sciences, social sciences, and computer and information science. In general terms, privacy can be defined as the right to be left alone, and the freedom from interference or intrusion [22].

There are three (3) main dimensions from which the privacy is described and analyzed: legal, social, and technical perspectives [23].

1) **Legal dimension**: privacy is a right that needs to be protected by laws, regulations, and policies [24], [25]. It is defined as a set of policies that enforce the protection of private information [26].
2) **Social dimension**: privacy depends on the behavior and the interactions of individuals as they conduct their daily affairs [23]. Rachels [27] stated that "privacy is necessary if we are to maintain the variety of social relationships with other people that we want to have". In such a way, privacy proves to be evolving with cultural changes and technological advances.
3) **Technical dimension**: this dimension aims to protect privacy through technical specifications by controlling (automatically and/or manually) data and information. Bunnig and Cap [28] described privacy as protecting personal information from malicious and unauthorized entities. Another definition of privacy is to hide some details from others [29]. The technical dimension is concerned with how the legal and social understandings of privacy can be represented and implemented in systems.

1) DEFINING PRIVACY IN THE CONTEXT OF MOSNs

We set the concept of privacy in OSNs by embracing the previous definitions in the context of MOSNs and defining privacy as the right to:

- decide what information to publish and who can access it.
- manage data (edit, download, and delete data at any point in time).
- assess the use, the processing, and the dissemination of data (who, how, what).
- give consent before any collection, use, or disclosure of data.
- specify the minimum acceptable of data in the system to provide the services.
- choose where to store data (locally or using the system's servers).
- protect all data available on the system from inquisitive entities.

Our definition of privacy in MOSNs captures, in essence, the idea of providing the users with the guarantee of data protection and with the autonomy (i) to assess the use and the access to data, (ii) to take appropriate action to protect their privacy, and (iii) to be assured that privacy is enforced and protected by the system.

C. PRIVACY AND SECURITY REQUIREMENTS

Until recently, privacy and security were treated as two separable concepts, and sometimes even opposite notions (privacy is breached in order to achieve security). However, the two concepts have a complementary and interdependent relationship. For example, privacy relies on access controls to protect sensitive data. Security of information refers to the protection of all types of information and systems [30], while information privacy ensures the fair protection of personal data.

The traditional considerations of information security have come up with a set of three security protection requirements called AIC-triad [31]–[33]:

1) **Availability** ensures access to authorized data and resources at any time and from everywhere.
2) **Integrity** ensures the reliability of the data, stored or in transit, and guarantees that any unauthorized modification is blocked.
3) **Confidentiality** protects the data content and prevents any unauthorized disclosure.

These three requirements are universally accepted and considered crucial to evaluate the security status in any IT system. They are used to achieve the desired level of protection and to assess the risks and investigate potential threats.

Unlike information security requirements, information privacy requirements are still new and not yet common or well established. A set of privacy-specific protection goals has been proposed by the National Institute of Standards and Technology (NIST) [34]. It represents the most important requirements from a privacy perspective. NIST has defined also three (3) privacy requirements in addition to the 3 security requirements (AIC) [35], namely:

1) **Predictability** aims at building trust and accountability between the system and the users about how the privacy-related data are processed. Predictability can be met with technical solutions such as logging and reporting.
2) **Manageability** provides the capacity to administer personal information, including editing and deleting information.

3) **Disassociability** aims at protecting an individual's identity and associated activities from disclosure and ensures that processing of personal information is dissociated to individuals.

The AIC-triad of information security along with privacy requirements provide a common model for addressing the legal, technical, and social dimensions of privacy and in accordance with our definition of privacy as explained in the previous section II-B.

III. RELATED WORK: PRIVACY EVALUATION AND SCORING IN OSNS

With the great success and spread of OSNs, there has been increasing research interest in mechanisms and methods that advocate for privacy protection. Many research studies have been performed on privacy preservation in OSNs by integrating new privacy protection mechanisms in already existing social networks or by proposing new built-in privacy systems. The increasing number of solutions and systems that aim to protect privacy in OSNs has led to the necessity to evaluate the efficiency and effectiveness of the proposed privacy protection mechanisms. However, evaluating privacy in a system is challenging since privacy itself is subjective and it is not easy to define as explained in section II-B.

One of the well-known processes to assess and evaluate privacy risk in systems is by conducting privacy impact assessment (PIA). PIA is part of Privacy by Design [36] and it is proposed to identify and quantify privacy risks in a system and take decisions on whether and how to mitigate, transfer, or accept these privacy risks [37]. In 2018, PIAs become mandatory for companies under article 35 in GDPR [38] (called Data Protection Impact Assessment-DPIA). PIAs are important tools for accountability and to ensure compliance with the regulation. However, PIAs have general purposes for all systems and they are not specific to OSNs, furthermore, the results are intended only for internal use by data controllers to ensure the implementation of appropriate measures and to demonstrate compliance with the applicable laws and regulations.

One of the first attempts to design a privacy metric for online social networks was proposed by Maximilien *et al.* [39] in 2009. The authors have proposed a framework to calculate the privacy score based on the sensitivity β_i and the visibility $V(i,j)$ of profile items $i \in i, \ldots, n$ of user j in a social network.

$$PR(i) = \sum_i PR(i,j) = \sum_i \beta_i * V(i,j)$$

However, the authors did not offer any dataset to measure the effectiveness of their model. This approach was extended in [40], where Liu and Terzi proposed a mathematical model based on the sensitivity and the visibility of profile attributes, using Item Response Theory (IRT) concept [41]. They evaluated the effectiveness of the score using synthetic and real-world datasets. However, the proposed model considered that the attributes are independent and it did not take into consideration the inferred data. Srivastava and Geethakumari [42] extended the model to include the hidden data. They proposed a privacy leakage score to quantify the level of privacy exposure from a message. Both models [40] and [42] assumed that the sensitivity and visibility are the same across all users. Petkos *et al.* [43] proposed a framework called PScore as a solution to the previous issues. PScore took into consideration the inferred information and it gave the users the possibility to change the scoring of attributes based on their personal preferences. Another privacy assessment framework based on the model of Liu and Terzi [40] was proposed by Pensa and Blasi [44]. The proposed score was based on both the sensibility and the visibility of user profile attributes, and it considered that the willingness ratio of a user to disclose information is proportional to the number of friends he or she has. The score measured the privacy leakage score and set a threshold for each user that when it is exceeded the user is notified.

Other research evaluated privacy from another aspect than sensitivity and visibility. PrivAware [45], proposed in 2009, quantified the privacy risk based on the amount of information inferred and set recommended actions to the users to enhance their privacy. Ngoc *et al.* [46] proposed a privacy metric based on the probability and entropy theory. It was based on the idea of how much of the hidden sensitive information an intruder can reveal from user's posts. Talukder *et al.* [47] proposed Privometer. It took into consideration the leakage of sensitive information from the applications used in the friends' profiles. Privometer ranked the relationships of users based on the amount of information leakage and proposed self-sanitizing recommendations to control the leakage. Another approach to measuring the privacy score was proposed by Akcora *et al.* [48]. They suggested computing the risk score based on the feedback from users about the friends' attitude and the similarities with the users. To evaluate the effectiveness of the model, the authors used Facebook and real datasets. Similarly, Vidyalakshmi *et al.* [49] proposed a privacy scoring framework based on the output of friends, their ranking and the total number of friends. The framework aimed at helping the users to assess the information sharing behavior and to take decisions of who can see what information.

In 2013, Nepali and Wang [50] introduced a new model SONET to monitor privacy exposure in real-time and to protect users from sensitive information disclosure. The privacy risk indicator (PIDX) was calculated based on the sensitivity and the visibility of attributes. SONET considered 2 aspects: attribute to attribute links (actor model) and user to user relationships (community model) and it included hidden information. The authors extended the actor model of SONET in [51]. They included 3 metrics: (1) known attribute list (direct, hidden and virtual), (2) attribute sensitivity and (3) attribute visibility. They introduced an implementation of SONET model called OSNPIDX tool in [52].

Table 1 summarizes the reviewed privacy scoring approaches.

TABLE 1. Overview of the reviewed privacy scoring approaches.

Privacy score	Description	Features
Privacy Scores [39], [40], [42]–[44]	A score generated based on the sensitivity and visibility of the items posted by an OSN user. Some scoring frameworks take into consideration the hidden and inferred information and the social graph.	Profile items, sensitivity per item, visibility per item. For some articles, leakage per item is also included.
PrivAware [45]	A score is calculated based on the total visible attributes divided by the total attributes in a profile.	The amount of information inferred in social networks
Privometer [47]	A score generated based on the sensitivity of the profiles of users and their social graphs.	Sensitive attribute inference from the information available in immediate friends' profiles.
Privacy score from social graphs [48], [49]	A score calculated based on the risk level in terms of the friends' attitude and the similarities with the users.	How much an attacker can reveal of relationships
Privacy Index [50]–[52]	A real-time score calculated based on the sensitivity and the visibility of public attributes.	Sensitivity score per item, visibility level per item.

IV. PROBLEM DEFINITION

The existing literature in the field of privacy scoring and evaluation in OSNs reveals some limitations to accurately calculate privacy scores:

- all the models evaluate the privacy in OSNs from the user's perspective. They assess privacy based on the visibility and the sensitivity of the attributes from a user's profile, such as name, age, address, phone number, etc. However, systems vary regarding the attributes they require users to provide [53], [54], hence the usage of attributes limits the evaluation of privacy in OSNs.
- the models consider the impact of the visibility of attributes to other users but not to the system provider. They do not include other aspects of privacy protection, for example, how the storage type can affect the privacy of the data.
- they do not consider security requirements in their evaluation, however, privacy and security concepts should not be separated since they are intertwined.
- the previous privacy metrics are different and they use system-specific measures, thus it is difficult to compare privacy scores across systems.

As a result, the present work proposes an enhancing privacy scoring model to quantify, assess and evaluate privacy protection in OSNs, with a focus on microblogging systems. The framework is systematic, generic, and universal. It allows comparing the privacy level across different systems. The proposed framework is designed with these considerations in mind:

- It must be universal and applicable to different MOSNs. Also, it should be fine-tuned to meet the needs of a specific situation or a system.
- It must take into consideration all the MOSNs stakeholders, namely: the users, the system operator, the third-party components, and the general public.
- It must include security requirements.
- It must take into consideration all types of data as explained in II-A.

V. PRIVACY METRICS DEVELOPMENT APPROACH: METHODOLOGY

The proposed framework is designed to guide end users to understand the impact of using social services on their privacy. It presents an instrument to measure the privacy level of an MOSN and to compare between different systems. The proposed framework also can be useful for system developers to assess the privacy controls implemented and to have recommendations to enhance the privacy of their systems.

The framework is based on a four-step methodology following the Plan-Do-Study-Act cycle (PDSA) [55]. The output is aimed for decision making and for assessing, monitoring, and predicting potential privacy threats in the system under investigation (SUI). It will enable the system providers to gauge how well their system is meeting the privacy and security objectives, and how well it protects the private data of the users. Fig. 1 summarizes the flow of the proposed methodology.

- **Step 1 (Plan):** the proposed framework begins by identifying the system under investigation and setting the scope and the depth of the assessment.
- **Step 2 (Do):** this step is concerned with gathering information about the SUI. It starts with a system architecture analysis to identify and understand the system structure, the business processes, the internal and external environment, the key assets and services, the security boundaries, and the implemented controls. Data can be gathered from different sources, including user feedback, risk assessment reports, research surveys, event loggers, etc.
- **Step 3 (Study):** the proposed framework, then, computes an overall privacy score based on the assessment of the impact of information security requirements (availability, integrity, and confidentiality) in addition to the privacy requirements as defined by NIST (predictability, manageability and disassociability). The framework answers 4 goals.
  - Goal 1: How the system protects itself from the privacy and security point of view.
  - Goal 2: How the privacy and the security of data are handled in the system.
  - Goal 3: How the system protects the users and the data.
  - Goal 4: How various assumptions and functionalities provided by the system might affect privacy and security.

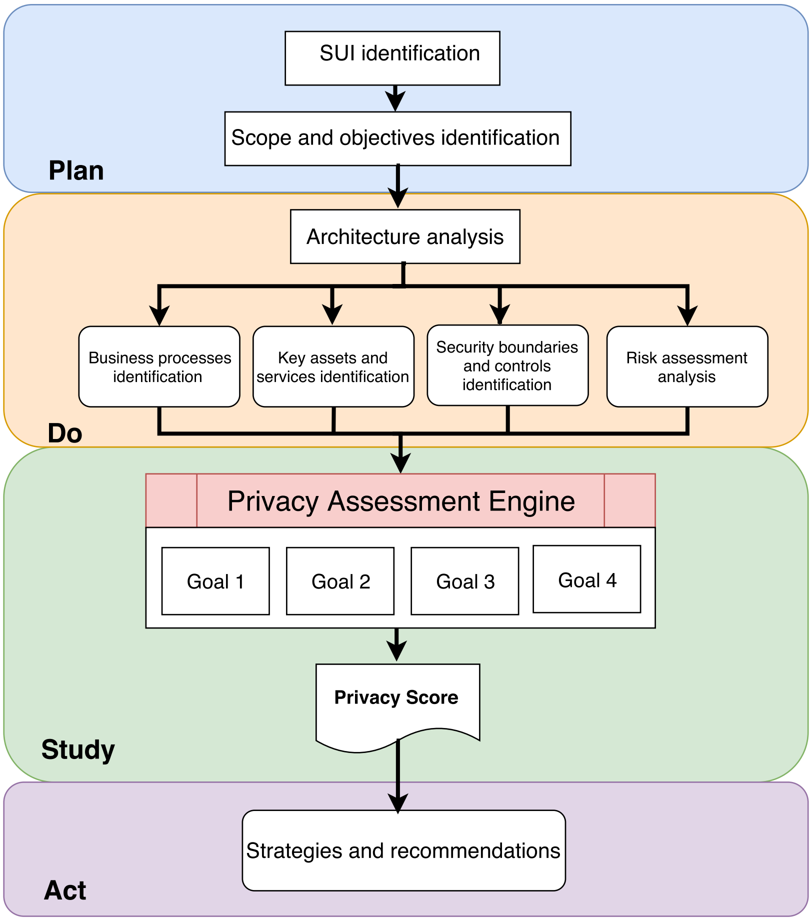
SUI identification
Scope and objectives identification
Plan
Architecture analysis
Business processes identification
Key assets and services identification
Security boundaries and controls identification
Risk assessment analysis
Do
Privacy Assessment Engine
Goal 1
Goal 2
Goal 3
Goal 4
Privacy Score
Study
Strategies and recommendations
Act

FIGURE 1. Framework methodology process.

- **Step 4 (Act):** based on the obtained score from step 3, the framework offers suggestions and recommendations for effectively controlling and protecting the privacy of the system.

During the assessment, it can be iterated back to step 2 at any time if the information gathered is not sufficient.

VI. INFORMATION PRIVACY ASSESSMENT ENGINE

Once the SUI is identified and analyzed and the data is gathered, the next step is to compute the privacy score. It should be noted that this research does not include methods and technologies to extract and gather information. With these operational characterizations in mind, the privacy assessment engine is developed following a Goal-Question-Metric (GQM) paradigm [56], [57]. GQM is a top-down derivation and a bottom-up interpretation approach that defines the relationship between the goals and the metrics as presented in Fig. 2.

In the GQM model, several questions are defined in such a way the metrics quantitatively answer each question, which leads to the achievement of the defined goal.

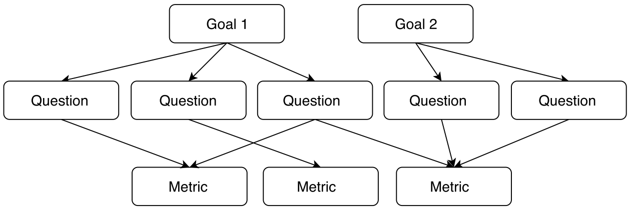

FIGURE 2. Goal-Question-Metric paradigm.

A. GOALS

Following the GQM model, the goals of the proposed framework are to analyze the SUI for the purpose of assessing and evaluating the level of privacy of:

1) how the system provider protects the users.
2) how the system provider protects the data.
3) how the system provider protects itself.
4) how various assumptions and functions in the system might affect the privacy and the security of the system.

B. ASSESSMENTS QUESTIONS

The assessment questions (AQs) in the framework can be answered either as Yes, No, or Not Addressed (N/A).

The questions are designed to cover and counter the known privacy threats and issues in OSNs [58]–[61] and they are derived from the set of criteria for classification and comparison discussed in [62]: (1) the type of service provided, (2) the architecture, (3) the storage and replication techniques, (4) the encryption mechanisms and key management, (5) the security goals, (6) the privacy goals, and (7) the functionalities. Furthermore, they are refined from the proposed requirements to protect privacy and security in social networks [19], [63]–[67]. They are also derived in compliance with our definition of privacy as discussed in section II-B (refer to appendix A for assessment questions).

Since the privacy scores resulted from the framework are used to compare different system, the assessment questions (AQs) are divided into two categories:

1) **Common AQs**: all assessment questions are common and generic to all systems. They are based on the definition of OSNs (refer to section II-A) and they cover OSNs stakeholders and data type as defined in section II-A. All the systems should answer the common questions and the scores resulted from this set are used to compare different systems.
2) **Specific AQs**: the questions are not common to all systems but specific to only some systems. For example, some systems provide direct messaging while others do not, or some systems provide the functionality of groups while others do not. The questions have an influence on privacy (either positively or negatively).

Furthermore, considering that the framework computes an overall privacy score based on the privacy risk present in the system and the privacy protection provided, the assessment questions are formulated in two different manners:

1) **Positive questions (Pos)** are oriented towards privacy protection.
2) **Negative questions (Neg)** have a negative impact on privacy (risk).

C. METRICS: THEORETICAL CALCULATION

Once each assessment question AQ_i in the common and specific sets is answered and the result is stored in $AnswerAQ_i$, the framework computes a privacy score from the responses. Table 2 presents the notations used in this paper.

The privacy score obtained from the proposed framework quantifies the level of the privacy protection provided in a system, under consideration of the existing privacy risk when using the services of the system. In other words, the overall score of an SUI is calculated based on the privacy risk of disclosed information and the privacy protection provided by the system. It is calculated as expressed in (1) and it is applicable to both common and specific scores:

$$TPS = \frac{PPS - PRS}{N} \tag{1}$$

where N depends if TPS is common score or specific score.

- In case of the common score, all questions are mandatory and should be counted including the questions answered as "N/A". Hence, N is expressed as (2)

$$N = N_{Common} = N_{PP} + N_{PR} + N_{NA} \tag{2}$$

- In case of the specific score, N includes only the questions answered as "Yes" or "No", whereas questions answered as "N/A" are ignored, as it is expressed in (3).

$$N = N_{Specific} = N_{PP} + N_{PR} \tag{3}$$

TABLE 2. Notation.

Notation	Description
TPS	Total Privacy Score
PPS	Privacy Protection Score
PRS	Privacy Risk Score
N	Total number of questions applicable to the SUI
N_{Common}	Number of answered questions in case of common set
$N_{Specific}$	Number of answered questions in case of specific set
N_{PP}	Number of answered privacy protection questions
N_{PR}	Number of answered privacy risk questions
N_{NA}	Number of answered N/A questions
$ScoreAQ$	Privacy score calculated for a question
Imp_{Priv}	Privacy Impact score
Imp_{Sec}	Security Impact score
AV	Accessibility Value
$Diff$	Data extraction difficulty

To compute the privacy score PPS and risk score PRS, the privacy assessment engine proceeds as follows:

1) Detect the category of the question (Common / Specific)
2) Detect the type of the question (Pos / Neg)
3) Compute $ScoreAQ$.
   - **Option1. In case of common score**
     $ScoreAQ$ is added to the privacy score PPS if positive questions have positive answers and negative questions have negative or N/A answers. Otherwise, $ScoreAQ$ is added to the risk score PRS if positive questions have negative or N/A answers and negative questions have positive answers. Algorithm 1 explains how privacy protection and privacy risk scores are computed in the case of common questions.
   - **Option2. In case of specific score**
     $ScoreAQ$ is added to the privacy score PPS if positive questions have positive answers and negative questions have negative answers. Otherwise, $ScoreAQ$ is added to the risk score PRS both positive questions have negative answers and negative questions have positive answers. The questions answered as N/A are ignored. Algorithm 2 explains how privacy protection and privacy risk scores are computed in the case of specific questions.

Figure 3 summarizes the process.

1) CALCULATION OF ScoreAQ

To compute PPS and PRS, we calculate a $ScoreAQ$ that differs for each AQ. $ScoreAQ$ measures the impact of the question in terms of privacy, security and visibility. It is calculated as

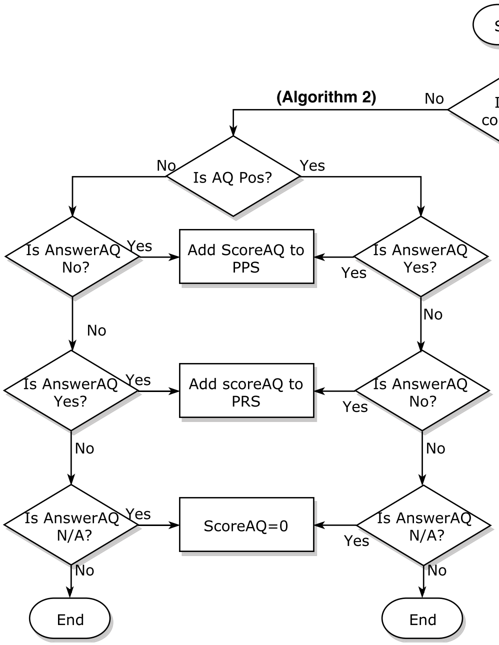

FIGURE 3. Privacy assessment engine workflow.

Algorithm 1 Algorithm for Computing *PPS* and *PRS* in Case of Common Score

```
$i \leftarrow 1$ while $i \leq numberofquestions$ do
    if $(AQ_i$ is $Pos$ && $AnswerAQ_i = Yes) \,||\, (AQ_i$ is $Neg$ && $AnswerAQ_i = No) \,||\, (AQ_i$ is $Neg$ && $AnswerAQ_i = N/A)$ then
        $PPS = PPS + ScoreAQ_i$;
    else if $AQ_i$ is $Neg$ && $AnswerAQ_i = Yes) \,||\, (AQ_i$ is $Pos$ && $AnswerAQ_i = No) \,||\, (AQ_i$ is $Pos$ && $AnswerAQ_i = N/A)$ then
        $PRS = PRS + ScoreAQ_i$;
end
```

Algorithm 2 Algorithm for Computing *PPS* and *PRS* in Case of Specific Score

```
$i \leftarrow 1$ while $i \leq numberofquestions$ do
    if $(AQ_i$ is $Pos$ && $AnswerAQ_i = Yes) \,||\, (AQ_i$ is $Neg$ && $AnswerAQ_i = No)$ then
        $PPS = PPS + ScoreAQ_i$;
    else if $AQ_i$ is $Neg$ && $AnswerAQ_i = Yes) \,||\, (AQ_i$ is $Pos$ && $AnswerAQ_i = No)$ then
        $PRS = PRS + ScoreAQ_i$;
    else if $AnswerAQ_i = N/A)$ then
        $ScoreAQ_i = 0$ (Question is ignored)
end
```

a function of the three factors: (i) Privacy impact, (ii) Security impact, and (iii) Visibility as shown in (4).

$$ScoreAQ = f(Imp_{Priv}, Imp_{Sec}, Visibility) \tag{4}$$

The formula to calculate *ScoreAQ* is based on the privacy score proposed in previous studies [40] and [42] (sensitivity of items multiplied by their visibility). However, in this proposed framework, instead of the sensitivity, we calculate the privacy and security impact of the items. This change was because the impact is a wider concept than sensitivity, while this later is a property or component of the impact [37].

To calculate the values of the factors to obtain *ScoreAQ*, different sources can be used [68]–[72].

a: CALCULATION OF PRIVACY IMPACT

Privacy Impact score Imp_{Priv} reflects the impact of the assessment question on the privacy. It is calculated as shown in (5):

$$Imp_{Priv} = Imp_{Pred} + Imp_{Manage} + Imp_{Diss}. \tag{5}$$

where:

- Imp_{Pred}: measures the impact of the question on predictability.
- Imp_{Manage}: measures the impact of the question on manageability.
- Imp_{Diss}: measures the impact of the question on disassociability.

The possible values of Imp_{Pred}, Imp_{Pred}, and Imp_{Diss} are between 0 and 5 as shown in table 3. Thus, the value of privacy impact is between 0 and 15.

TABLE 3. Impact value for privacy and security.

Values	Numerical Value
Very High	5
high	4
Moderate	3
Low	2
Insignificant	1
None	0

TABLE 4. Accessibility value.

Values	AV in case of PPS	AV in case of PRS
Publicly available	1	5
Accessible by friends of friends	2	4
Accessible by system provider	3	3
Accessible by friends	4	2
Not accessible except data owner	5	1
None	0	0

b: CALCULATION OF SECURITY IMPACT

Security Impact score Imp_{Sec} reflects the impact of the assessment question on the security. It is calculated as shown in (6):

$$Imp_{Sec} = Imp_{Avail} + Imp_{Conf} + Imp_{Integ}. \tag{6}$$

where:

- Imp_{Avail}: measures the impact of the question on availability.
- Imp_{Conf}: measures the impact of the question on confidentiality.
- Imp_{Integ}: measures the impact of the question on integrity.

The possible values of Imp_{Avail}, Imp_{Conf}, and Imp_{Integ} are between 0 and 5 as shown in table 3. Given these ranges of values, the final security impact has a score between 0 and 15.

c: CALCULATION OF VISIBILITY

The visibility score determines how accessible the data discussed in the question. For calculating the visibility, we consider two factors: accessibility and data extraction difficulty. For visibility score, we are using the formulas proposed by Aghasian *et al.* [73]. The visibility score is calculated as shown in (7).

$$Visibility = AV * Diff. \tag{7}$$

- **Accessibility Value** (AV) measures the permissions given to share information with others. In other words, it indicates how many people can access the shared information. The accessibility value depends on whether the score calculated is PPS or PRS. Table 4 shows the possible accessibility values.
- **Data extraction difficulty** ($Diff$) measures the difficulty of extracting private information from the formats of data discussed in the assessment question. For calculation of difficulty, 4 levels have been defined from 0 to 3. The extraction difficulty value depends also on whether the score calculated is PPS or PRS.

TABLE 5. Data extraction difficulty values.

Values	$Diff$ in case of PPS	$Diff$ in case of PRS
Low difficulty,	1	3
Medium difficulty	2	2
High difficulty	3	1
None	0	0

Table 5 shows the possible values for extraction difficulty. Naturally, the more accessible an item is, the less difficult is to extract data.

The value of visibility is between 0 and 15.

d: CALCULATIONS OF ASSESSMENT QUESTION SCORE ScoreAQ

As expressed in (4), $ScoreAQ$ depends on three factors (i) privacy impact Imp_{Priv}, (ii) security impact Imp_{Sec}, and (iii) visibility $Visibility$. It is not necessary that all factors are applied in calculating $ScoreAQ$. The assessment questions can have an impact only on one factor (for example only privacy is applicable), or on two factors (for example privacy and visibility are applicable), or on all three factors (privacy, security, and visibility). Thus, there are 3 cases to calculate $ScoreAQ$.

1) Case A: only one factor is applicable (not null)

$$ScoreAQ = \frac{20}{3} * Factor. \tag{8}$$

2) Case B: 2 factors are applicable (not null)

$$ScoreAQ = \frac{4}{9} * Factor1 * Factor2. \tag{9}$$

3) Case C: all factors are applicable (not null)

$$ScoreAQ = \frac{4}{135} * \prod_{i=1}^{3} Factor_i. \tag{10}$$

The possible values of $ScoreAQ$ are between 0 and 100. The purpose of the weight added to the equations is to harmonize the score between all the case since the maximum value possible for each factor is 15.

2) SCORE EVALUATION AND COMPARISON

To understand the overall privacy score obtained in the system under investigation, we compare it against two systems of reference (an ideal system and a flawed system). The systems of reference are hypothetical microblogging systems where the ideal system is 100% efficient in protecting the privacy of users and data, while the flawed system is 100% deficient in protecting privacy. In other words, the total score of the ideal system is based only on PPS ($PRS = 0$), as shown in (11), and the total score of the flawed system is based only on PRS ($PPS = 0$), as shown in (12). The total privacy score obtained of every investigated system is between TPS_{Flawed} and TPS_{Ideal} (see (13)).

$$TPS_{Ideal} = \frac{PPS}{N}. \tag{11}$$

$$TPS_{Flawed} = \frac{-PRS}{N}. \tag{12}$$

$$TPS_{Flawed} \leq TPS_{SUI} \leq TPS_{Ideal}. \tag{13}$$

To understand further the obtained overall privacy score and to facilitate the comparison between different systems, we convert the values *TPS* into a percentage (%) ranging between the flawed system (TPS_{Flawed} would represent 0%) and the ideal system (which would represent 100%).

TABLE 6. Example to calculate the accuracy.

	Answers in Ideal	Answers in SUI	Outcome
AQ1	Yes	Yes	1
AQ2	Yes	No	0
AQ3	No	Yes	0
AQ4	No	No	1

Another type of score evaluation is to compute the accuracy of the SUI. To do so, we compare the SUI against the ideal system based on the answers of the assessment questions for each system (ideal and SUI). We calculate the number of true results where the answers of SUI match the answers from the ideal system (true positives (TP) and true negatives (TN)), as shown in the example in table 6. The accuracy of SUI then is computed as the number of questions that match the answers from the ideal system among the total number of assessment questions, as shown in (14). Naturally, the accuracy of the ideal system is 100% and the accuracy of the flawed system is 0% (see (15) and (16)), while the accuracy of the SUI is between 0% and 100% as shown in (17).

$$Accuracy = \frac{Numberofcorrectlyansweredquestions}{N}. \tag{14}$$

$$Accuracy_{Ideal} = 100\% \tag{15}$$

$$Accuracy_{Flawed} = 0\% \tag{16}$$

$$Accuracy_{Flawed} \leq Accuracy_{SUI} \leq Accuracy_{Ideal} \tag{17}$$

TPS_{SUI} gives a general overview of the score of privacy level provided in the system under investigation. It can be used by users to compare between different systems and choose the most adequate system based on their necessity. While the accuracy measures to what extent the system under investigation conforms to a standard and how close it is to an ideal system that protects the privacy of users. The result obtained from computing the accuracy can help the developers to compare their system with an ideal one and to evaluate and fix the privacy flaws in the system.

VII. EVALUATION AND EXPERIMENTS: CASE STUDIES

To prove the applicability of our framework, we compare five known online social networks that offer microblogging services. First, we present the results of applying our framework and then we compare the systems using our model.

A. STEP 1: SCOPE AND OBJECTIVES DEFINITION

The proposed framework is used to assess and evaluate the level of privacy of Facebook [74], Twitter [75], Tumblr [76], Diaspora [77], and GAP [78]. The choice of Facebook, Twitter, and Tumblr was due to their popularity with 2.9 billion monthly active users combined [79]. While the choice of Diaspora and GAP was due that both of the platforms are designed and built with privacy protection in mind. The purpose is to assess the privacy protection mechanisms in each system and to monitor the status of the level of privacy provided. The proposed framework will facilitate the comparison between different systems and it will be useful for the improvement and enhancement of the privacy protection by applying corrective actions, based on the observed results.

B. STEP 2: SUI ANALYSIS AND DATA GATHERING

The second step starts with analyzing the architecture of systems and the functionalities provided. The data gathered is summarized in the following:

1) FACEBOOK

Facebook [74] is one of the well-known online social networks offering its users a platform where they can share their daily life with friends and connections.

Facebook uses a centralized architecture with MySQL database infrastructure and Global Transaction ID with MySQL semi-synchronous replication [80], [81]. It uses Transport Layer Security (TLS) to secure the communication between the clients and the servers.

Users can create accounts, add, accept or decline friendship requests, search of a user, post messages, reshare, and comment on others' posts. They can mention other friends and they can post directly on their friends' walls, provided that they are authorized to do so. Facebook displays recommendation based on the location and interests of users and it gives the possibility to the users to search for a user, a page or a group. Unless the user restricted the access, Facebook's profile is by default public and anyone on the Internet can access it and see what is shared and the relationship between the users.

2) TWITTER

Twitter is an online social network that provides microblogging services. It allows users to post, share, and comment short messages called "tweets". The number of maximum character was expanded from 140 to 280 characters [82]. Twitter has a centralized infrastructure with a next-generation distributed database to provide reliability, availability, and low latency [83]. It uses TLS to secure communication between the clients and the servers. Twitter provides privacy policies and it retains the right to access and analyze the data stored to recommend new contents or ban abusive users and offensive hashtags [84].

Twitter allows creating a profile using personal information (e.g. full name, phone number). It provides an optional service to verify the account to authenticate the identity of the person or company that owns the account [85]. The users can post text, photos, videos, location information, etc. The profile, tweets and the list of followers are public by default and accessible to all the internet, but the users have the option to restrict the accessibility of the profile and the tweets to only their followers. However, some information provided in the profile is always public like biography and picture. Twitter has one way following model where anyone can read tweets or mention other users although they do not follow them. The interface of Twitter displays a list of trending hashtags (#) and topics along with recommendations of potential new profiles to follow. Twitter also provides direct messaging to communicate privately with followers.

3) Tumblr

Tumblr [76] is part of Oath Inc. [86] and it has more than 400 million active users by April 2019 [87]. The platform uses a centralized architecture with Redis, HBase and MySQL databases [88] and Multi-source Replication from MariaDB [89] to protect the availability of their services. It uses TLS to secure communication between the clients and the servers.

Users can post messages, comment or share others' posts, tag interests, mention or search other users. The profiles are by default visible to all Internet, but the users restrict the accessibility of their blog to only viewed by the users of Tumblr.com [90]. Yet, the profile and all the posts shared on the blog are visible to the other Tumblr users even if they are not in the followers' list.

4) DIASPORA

Diaspora is a user-owned microblogging OSN based on the free Diaspora software [91]. It has a federated architecture allowing users to host their accounts by setting up their own server/pod. Users have also the possibility to create their profiles on an existing pod that is open or upon receiving an invitation from the owner of a closed pod. They can choose a pod based on the frequency of updating software version or the ratings of the pod. Diaspora network backs up the data on a different server or on a different hard drive. Diaspora uses Pretty Good Privacy (PGP) where a unique public/private key pair and an ID are assigned to every user created on a pod. The pod is the one in charge of encrypting and decrypting requests before passed to users. The administrators of pods have read and write access rights to the clear data stored on their pods [92].

Profiles in Diaspora have two parts: (1) public that is visible to all logged-in users (e.g. name, interests, and photo), and (2) private with detailed information and accessible to only authorized users (e.g. biography, location, gender, and birthday). Users can publish posts either (1) publicly where any logged-in user can read, comment on, reshare, and like the posts, or (2) privately where only authorized followers can read, comment on, and like but not reshare the private posts. Diaspora offers also instant messaging services called conversations where users can send private messages. In Diaspora, it is possible to follow others and read their posts without mutual friending them [93], [94].

5) GAP.ai

GAP.ai [78] is a microblogging system that mimics the functionalities of Twitter while offering its users the freedom of speech and thought. GAP service retains the right to store and administer users' data. It banned the first GAP user in January 2017 [95]. Gab comes in two versions: the free GAP and GAPPro. GAPPro is a paid version that allows users to use private group chats, and to go live [96].

The users in GAP.ai can make their profiles public or private and add new followers. They can send, quote a post, tag an interest, mention another user, search for other users and interests, and get recommendations of potential friends and hot topics. The posts and the lists of followers are public and visible to any user of Gab. Traffic between clients and servers is encrypted using TLS to secure the traffic between the clients and the servers.

Currently, Gab uses a centralized architecture. The developers have announced that they will change the architecture to a decentralized architecture to build a true censorship-resistant and community-powered system [97].

C. STEP 3: PRIVACY SCORE COMPUTATION

The purpose of this experiment is to evaluate the privacy level provided in five known microblogging OSNs.

In this study, we derived the assessment questions for common and specific scores. In total, we had 48 common assessment questions and 68 system-specific assessment questions (refer to appendix A). The questions are divided in the common set into 6 parts: (1) profile, (2) relationship, (3) posts, (4) data storage, (5) data collection, (6) data encryption. The common questions should all be answered with Yes, No, or N/A. The questions in the specific set are divided into 11 parts: (1) profile, (2) posts, (3) groups, (4) data collection, (5) data encryption, (6) functionalities, (7) architecture and application, (8) settings, (9) privacy policies, (8) feedback, and (9) API and third-party relationships. Each question can be answered with Yes or No while the questions answered as N/A are ignored ($ScoreAQ = 0$).

Once the questions are answered, we derived the values of answers from [68]–[72]. Then, we calculated the $ScoreAQ$ for each question for all the systems under investigation. (refer to appendix B for more details).

1) **Common Privacy Scores**

   Following algorithm 1, we calculated the privacy risk score PRS based on how much information is disclosed and the privacy protection score PPS based on how the privacy of users is protected in each investigated system. The results of PRS and PPS are presented in table 7 for the score of the reference systems, table 8 for the deployed systems

   Once the privacy risk score PRS and the privacy protection score PPS are computed for each investigated system, using (1), we calculated the common overall privacy score TPS for the reference systems and the investigated systems, as shown in tables 9 and 10.

TABLE 7. Common PPS and PRS for the reference systems.

	Flawed system	Ideal system
N_{PP}	0	48
PPS_{Common}	0	1475
N_{PR}	48	0
PRS_{Common}	-1314	0

TABLE 8. Common PPS and PRS for the investigated systems.

	Facebook	Twitter	Tumblr	Diaspora	Gab
N_{PP}	19	23	27	40	19
PPS_{Common}	497	524	630	978	600
N_{PR}	29	25	21	8	29
PRS_{Common}	-619	-716	-632	-279	-597

TABLE 9. Common TPS for the reference systems.

	Flawed	Ideal
N_{Common}	48	48
TPS_{Common}	-27,38	30,73
Percentage (%)	0%	100%

TABLE 10. Common TPS for the investigated systems.

	Facebook	Twitter	Tumblr	Diaspora	Gab
N_{Common}	48	48	48	48	48
TPS_{Common}	-2,54	-4,00	-0,04	14,56	0,06
Percentage(%)	43%	40%	47%	72%	47%

TABLE 11. Specific PPS and PRS for the reference systems.

	Flawed system	Ideal system
N_{PP}	0	58
$PPS_{Specific}$	0	1838
N_{PR}	62	0
$PRS_{Specific}$	-1766	0

TABLE 12. Specific PPS and PRS for the investigated systems.

	Facebook	Twitter	Tumblr	Diaspora	Gab
N_{PP}	32	34	31	33	27
$PPS_{Specific}$	856	983	928	889	838
N_{PR}	31	30	23	14	34
$PRS_{Specific}$	-882	-827	-532	-317	-959

2) **Specific Privacy Scores**
Similar to the common privacy score calculated in the previous section, we calculated the specific privacy scores for the systems under investigation. Using algorithm 2, we calculated the privacy risk score *PRS* and the privacy protection score *PPS* for each investigated system. The results of *PRS* and *PPS* are presented in table 11 for the score of the reference systems and table 12 for the investigated systems.
Once the privacy risk score *PRS* and the privacy protection score *PPS* are computed for each investigated system, using (1), we calculated the common overall privacy score *TPS* for the reference systems and the investigated systems, as shown in tables 13 and 14

3) **Accuracy Scores**
Also, we compared the accuracy of investigated systems with the ideal system by applying (14). We found that Facebook is the least accurate of the systems with only 40% accuracy comparing with other systems. While Diaspora has 83% accurate in the common set (refer to table 15 for more details). From the results of accuracy, we can observe that Diaspora is more conform to the ideal system followed by GAP, Tumblr, Twitter, then Facebook. In other words, Diaspora and GAP tend to implement privacy controls and use techniques to protect the privacy of users better than the other system, which is expected since both Diaspora and GAP are privacy-oriented.

TABLE 13. Specific TPS for the reference systems.

	Flawed	Ideal
$TPS_{Specific}$	62	58
$TPS_{Specific}$	-28.48	31.69
Percentage (%)	0%	100%

TABLE 14. Specific TPS for the investigated systems.

	Facebook	Twitter	Tumblr	Diaspora	Gab
$N_{Specific}$	63	64	54	47	61
$TPS_{Specific}$	-0.41	2.44	7.33	12.17	-1.98
Percentage (%)	47%	51%	60%	68%	44%

TABLE 15. Accuracy of the investigated systems.

	Facebook	Twitter	Tumblr	Diaspora	Gab
N	48	48	48	48	48
Nº correctly answered	19	23	27	40	29
Accuracy	40%	48%	56%	83%	60%

D. STEP 4: COMPARISON AND ANALYSIS

The goal of this experiment was to compare the privacy scores obtained in the investigated systems. Once the privacy scores are computed using the proposed algorithmic model, the framework offers recommendations and suggestions to enhance the obtained score and minimize the privacy risk score. The privacy assessment of the SUIs identifies the risk items that should be addressed by the system provider. It gives also instructions to users about the privacy settings and parameters that should be changed from the default status before starting using the services of the system investigated. Table 16 shows an example of recommendations provided by the framework in the case of the investigated systems.

The comparison between the systems uses the common privacy scores since all the SUI have to answer the same number of assessment question (48 in the case of this study). The obtained reference common values in this test model are -27.38 for the flawed system and 30.73 for the idea system.

By comparing the results in table 8, it can be seen that Diaspora has answered 40 questions as protecting the privacy and only 8 questions as risk generating, followed by Tumblr that has answered 27 privacy-protecting questions. Even though Facebook and Gab have answered 19 questions each as protecting privacy and 29 questions as risk generating, the difference between two systems is GAP gives more

TABLE 16. Example of recommendations.

Assessment Questions (indicative)	Imp_Priv	Imp_Sec	$Visibility$	$ScoreAQ$	Recommendations
Facebook					
1. Can the unregistered users access the service without creating a profile?	14	7	15	-44	Only the authorized users should be able to access the profile
2. Can the users change the visibility of their relationship list?	12	7	2	5	No recommendation
3. Is a new post visible to the public by default on the user's timeline?	12	7	15	-38	Post should be private by default and the users should be allowed to change the visibility of their posts
Twitter					
1. Can the unregistered users access the service without creating a profile?	14	7	15	-44	Only the authorized users should be able to access the profile
2. Can the users change the visibility of their relationship list?	12	7	15	-38	The users should be allowed to restrict the visibility of the relationship list
3. Is a new post visible to the public by default on the user's timeline?	12	7	15	-38	Post should be private by default and the users should be allowed to change the visibility of their posts
Tumbler					
1. Can the unregistered users access the service without creating a profile?	14	7	15	-44	Only the authorized users should be able to access the profile
2. Can the users change the visibility of their relationship list?	12	7	15	38	No recommendation
3. Is a new post visible to the public by default on the user's timeline?	12	7	15	-38	Post should be private by default and the users should be allowed to change the visibility of their posts
Diaspora					
1. Can the unregistered users access the service without creating a profile?	14	7	1	3	No recommendation
2. Can the users change the visibility of their relationship list?	12	7	1	3	No recommendation
3. Is a new post visible to the public by default on the user's timeline?	12	7	12	30	No recommendation
GAP					
1. Can the unregistered users access the service without creating a profile?	14	7	15	-44	Only the authorized users should be able to access the profile
2. Can the users change the visibility of their relationship list?	12	7	4	10	No recommendation
3. Is a new post visible to the public by default on the user's timeline?	12	7	15	-38	Post should be private by default and the users should be allowed to change the visibility of their posts

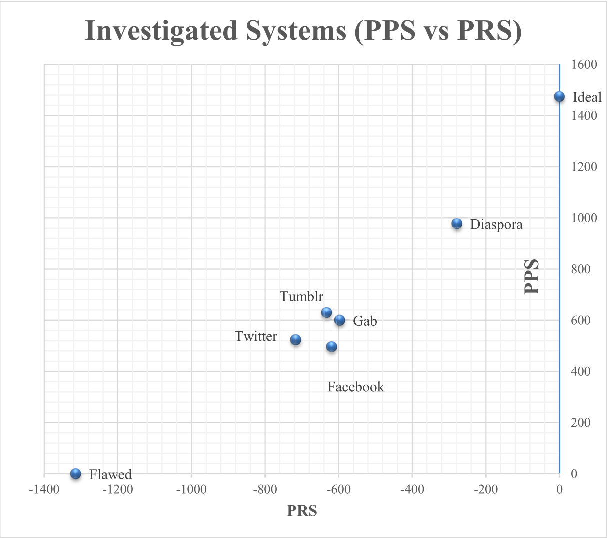

FIGURE 4. Comparison between scores PPS and PRS for the investigated systems.

access control over data in comparison with Facebook. Fig. 4 compares between PPS and PRS scores obtained for each system.

Applying (4) and algorithm 1, we found that the overall common privacy score of Twitter is -4.00 (40% in comparison with the ideal system). It has the lowest score obtained in comparison with other systems, while Diaspora obtained 14.56 (72% in comparison with the ideal system). The results showed as well that GAP and Tumblr are 47% in comparison with the ideal system, while Facebook got 43% in comparison with the ideal system.

The reason why Twitter got a lower privacy score comparing with Facebook is because of the access settings implemented for the users to control the visibility of their profiles, list of followers, and posts, for example, in Twitter, users cannot specify from whom to receive follow requests or change the visibility of the followers list contrary to Facebook. As for GAP, the developers claimed that they provide freedom of speech and thought, however, the platform has centralized architecture and the providers can access and administer the data of users. Fig. 5 and 6 show a graphical representation of the different results obtained for the investigated system.

VIII. DISCUSSION

The present framework methodology is similar to the methodology proposed by the French Data Protection

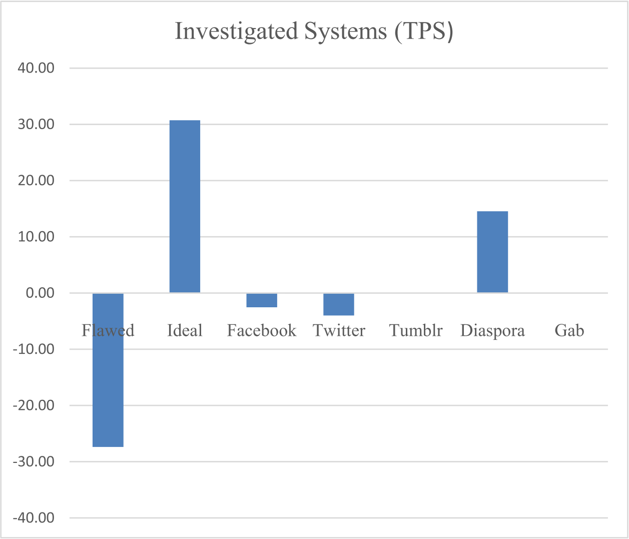

FIGURE 5. Comparison the investigated systems in terms of TPS.

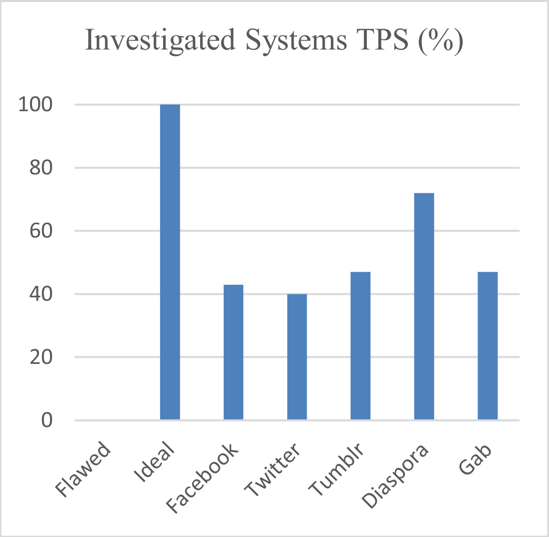

FIGURE 6. Comparison the investigated systems in terms of TPS (%).

Authority (CNIL) [98] to carry out PIA in the context of GDPR. The CNIL methodology is composed of 4 iterative steps: (1) study of the context to gain a clear overview of the data processing operations under consideration, (2) study of the fundamental principles and assessment of controls protecting data subjects' rights, (3) study of the risks related to the security of data (how much damage would be caused by all the potential impacts), and (4) validation of the PIA and decision whether or not to accept the results in light of the findings.

However, the CNIL methodology differs from the proposed framework in many regards. The CNIL methodology is not specific to online social networks and it is intended for internal use by system providers and data controllers, the results of the assessment are not shared with the end users but they are used to improve the implementation of privacy protection techniques and to demonstrate the compliance with GDPR. Moreover, it does not provide a privacy score and it considers only the calculation of potential privacy risks without considering the mechanisms already implemented to protect the privacy in the system.

The proposed framework differs significantly from existing privacy scores and privacy assessment processes. As discussed in section IV, previous models computed privacy scores in OSNs based on the sensitivity and the visibility of users' attributes (such as profile's items, relationships,...). Our approach is related to these models, the formula proposed to calculate the *ScoreAQ* is based on the visibility and we replaced the sensitivity by the impact of privacy and security as two additional factors in calculating privacy score, since the sensitivity is a property of the impact. Furthermore, the overall privacy score is not based only on the assessment risk available in the investigated system, but also on the privacy protection provided by the system to their users.

To the best of our knowledge, we are the first to present an algorithmic model that computes privacy score from this perspective and focuses on microblogging OSNs.

IX. CONCLUSION

Users enjoy sharing their activities and interests and connecting with friends using the services of OSNs. However, a large amount of personal information (sensitive or non-sensitive) is accessible to all Internet. Moreover, advanced data analytics have made it easy to extract hidden information. One of the challenges is to assess and evaluate the privacy protection techniques present in an OSN system.

In this paper, we have presented a comprehensive and generic framework to compute the privacy score in microblogging Online Social Networks. We presented a systematic methodology to develop privacy metrics in MOSNs based on the Plan-Do-Study-Act (PDSA) cycle. Then, we proposed a novel algorithmic model to compute the privacy score in MOSNs. The model is based on the impact of privacy and security requirements, accessibility, and extraction difficulty of information in MOSNs. The proposed framework can be used to obtain evidence of the effectiveness and efficiency of the privacy mechanisms implemented in the system under investigation. Furthermore, the framework aims to quantify, measure, and evaluate the privacy of the system under investigation and compare different MOSNs. We have demonstrated the feasibility of our framework using five real social networks, Facebook, Twitter, Tumblr, Diaspora, and GAP. The obtained results show the potential of the proposed framework.

This work represents the first steps towards developing a practical framework to quantify privacy in microblogging systems. To further refine the present work, we plan to extend the application of the proposed framework to a larger number of microblogging Online Social Networks. Also, we plan to explore further generalizations and improvements of the algorithms and to design a user interface for the framework. In addition, we intend to develop techniques to automatically extract values from the data for the parameters that our framework uses to compute privacy scores.

APPENDIX A ASSESSMENT QUESTIONS

A. COMMON SET

- **Profile**

1) Can the unregistered users access the service without creating a profile?
2) Are the following items required to create a profile?
   - Name
   - Surname
   - Username
   - National Identifying Number
   - Age
   - Birthdate
   - Gender
   - Avatar or profile image
   - Civil Status
   - Phone Number
   - Email address
   - Postal address
   - Personal interests and preferences
   - Profession or Education
   - Political affiliation
   - Religious affiliation
   - Sexual orientation
   - Contact information
3) Is it required to be authenticated to access the services?
4) Can the user create multiple profiles with the same credentials?
5) Once the profile is created, is it public by default?
6) Can the user change the visibility of the profile?
7) Can the user edit the items in the profile?
8) Can the user delete the profiles?
9) Can the SUI edit the user information without consent?
10) Can the SUI delete profiles without the consent of their owners?

- **Relationships**

11) Are the friendship requests accepted automatically without consent?
12) Can the user accept a friendship request?
13) Can the user deny a friendship request?
14) Can the user restrict from whom to receive friendship requests?
15) Is the relationship list public by default?
16) Can the user change the visibility of the relationship list?
17) Can the user delete friends from the friendship list?

- **Posts**

18) Is a new post visible to the public by default on the user's timeline?
19) Can the user specify the audience of a new post?
20) Can the user edit their posts?
21) Can the user delete their posts?
22) Can the SUI edit user's data without consent?
23) Can the SUI delete user's data without consent?

- **Data storage**
  - **Option 1.**
    24) Is data stored in centralized databases handled by the SUI?
    25) Does the SUI replicate the data stored?
    26) Does the SUI provider have a secondary site for data storage?
    27) Does the SUI delete the data stored once the users delete it?
  - **Option 2.**
    24) Is data stored in pods or clients' machines?
    25) Does the SUI duplicate and store the data stored in the pods?
    26) Is the data replicated?
    27) Can the SUI delete the data stored without consent?
  - **Option 3.**
    24) Is data stored in Hybrid servers?
    25) Does the SUI handle the storage of the data of users?
    26) Is the data replicated?
    27) Can the SUI delete the data stored without consent?

- **Data collection**

28) Does the SUI collect data from users? (If yes, go to Specific AQs)
29) Does the SUI analyze and sell data? (If yes, go to Specific AQs)

- **Data encryption**

30) Is the data stored encrypted? (If yes, go to Specific AQs)
31) Are the encrypted communication channel used when transferring data? (If yes, go to Specific AQs)

B. SPECIFIC SET

- **Profile**

1) Can the user reset their authentication?
2) Can the user recover their authentication?
3) Does the SUI enforce a limit of consecutive invalid login attempts?
4) Does the SUI send alerts for unknown logins?
5) Can the user deactivate their profiles without deleting them?

- **Posts**

6) Does the SUI add location information by default to the posts?
7) If yes, can the user deactivate the option of location information?

- **Groups**

8) Is the group once created visible to all by default?

9) Can the group administration restrict the visibility of the group?
10) Can the user be added to a group without their consent?
11) Can anyone add users to the group?
12) Can anyone delete users to the group?
13) Is the membership list public by default in the group?
14) Can the group administrator control the visibility of the membership list?
15) Can the user leave a group?
16) Can SUI delete a group without the consent of its owner?

- **Data collection**
  - Does the SUI collect data from users? (if yes continue) (Not counted)

17) Does the SUI explain the data collected, what type and what for?
18) Does the SUI ask for users' consent to collect data?
19) Can the user withdraw their consent?
20) Does the user control the data to be collected by the SUI?

  - Does the SUI analyze and sell data? (if yes continue) (Not counted)

21) Does the SUI ask users' permission to analyze and sell the data?
22) Does the SUI use anonymization mechanisms before publishing data?
23) Does the SUI specify to whom the data are sold?
24) Can the user specify to whom the data will be sold?

- **Data encryption**
  - Is the data stored encrypted? (if yes continue) (Not counted)

25) Is the encryption algorithm used weak?
26) Are the key lengths weak?
27) Does the SUI generate the keys?
28) Are the keys stored in the SUI?

  - Are the encrypted communication channel used when transferring data? (if yes continue) (Not counted)

29) Does the SUI use weak protocols for SSL/TLS?
30) Does the SUI use weak key length for SSL/TLS?

- **Functionalities**

31) Can the user post messages on others' timelines (non-friends)?
32) Can the user control who can post a message in their timeline?
33) Can the user mention/tag any other user (non-friends) in the comments?
34) Can the user control who can mention them in the comments?
35) Can the user share or resend friends' posts?
36) Can the user share or resend any other user's posts?
37) Can the user comment on any other users' posts?
38) Can the user receive direct messages from anyone?
39) Can the user restrict from whom to receive a DM?

TABLE 17. Common ScoreAQ for the investigated systems.

AQs	Facebook	Twitter	Tumblr	Diaspora	GAP
Profile					
AQ1	-44	-44	-44	3	-44
AQ2					
Name	-15	-15	15	15	-15
Surname	-16	-16	16	16	-16
Username	1	1	1	4	1
SSN	16	16	16	16	16
Age	-13	13	-8	13	13
Birthdate	-13	-13	13	13	13
Gender	-13	13	13	13	13
Avatar	-15	-15	15	15	15
C.Status	-6	15	15	15	15
Phone.N	-15	-15	-15	15	15
Email	1	1	1	4	15
Postal	-7	16	16	16	16
Interests	-5	13	-13	13	-13
Profession	-13	13	13	13	13
Politics	-7	16	16	16	16
Religion	-7	16	16	16	16
Sexual.Pr	-7	16	16	16	16
Contact	-6	15	15	15	15
AQ3	69	69	69	69	69
AQ4	64	64	64	64	64
AQ5	-38	-38	-38	-20	-38
AQ6	-38	20	-38	30	10
AQ7	18	18	18	18	18
AQ8	18	-18	18	18	18
AQ9	-13	-13	-13	22	5
AQ10	-13	-13	-13	22	-13
Relationships					
AQ11	38	-38	-38	-38	-38
AQ12	38	38	-38	-38	-38
AQ13	32	32	-32	-32	-32
AQ14	4	-18	-27	-27	-27
AQ15	-38	-38	-38	3	-38
AQ16	5	-38	38	3	10
AQ17	25	-25	25	25	25
Posts					
AQ18	-38	-38	-38	30	-38
AQ19	5	-38	38	30	-38
AQ20	44	-44	44	44	44
AQ21	44	44	44	44	44
AQ22	-23	-23	-23	8	-23
AQ23	-23	-23	-23	-23	-23
Storage	**Opt1**	**Opt1**	**Opt1**	**Opt2**	**Opt1**
AQ24	-30	-30	-30	56	-30
AQ25	11	11	11	56	11
AQ26	20	20	20	49	20
AQ27	-30	-30	-30	64	10
Data collection					
AQ28	-32	-32	-32	-32	-32
AQ29	-32	-32	-32	32	-32
Data encryption					
AQ30	-69	-69	-69	-69	-69
AQ31	44	44	44	44	44

40) Does the SUI provide users with updates, news and advertisement contents based on their behavior and interests?
41) Is yes, can the user accept or refuse to receive updates, news and advertisement contents?
42) Can the user search other users by an identifier (name, email,...)?
43) If yes, can the user control who can search for them by identifier?
44) If yes, can the user control who can search for them by email?

TABLE 18. Specific ScoreAQ for the investigated systems.

AQs	Facebook	Twitter	Tumblr	Diaspora	Gab
Profile					
AQ1	35	35	35	35	35
AQ2	29	29	29	29	29
AQ3	-23	-23	23	-23	-23
AQ4	23	23	23	-23	-23
AQ5	18	18	-18	-18	-18
Posts					
AQ6	8	-23	8	8	23
AQ7	0	38	0	0	0
Groups					
AQ8	-38	-38	0	0	-38
AQ9	10	30	0	0	-38
AQ10	-54	-54	0	0	54
AQ11	-43	-43	0	0	-43
AQ12	-43	-43	0	0	43
AQ13	-38	-38	0	0	-38
AQ14	8	-27	0	0	-27
AQ15	27	27	0	0	27
AQ16	-32	-32	0	0	-32
Data collection					
AQ17	32	32	32	32	32
AQ18	32	32	32	32	32
AQ19	-32	-32	32	-32	-32
AQ20	-32	-32	-32	-32	-32
AQ21	-35	35	-35	0	-35
AQ22	40	40	40	0	40
AQ23	35	-35	-35	0	-35
AQ24	-24	-24	-24	0	-24
Data encryption					
AQ25	0	0	0	0	0
AQ26	0	0	0	0	0
AQ27	0	0	0	0	0
AQ28	0	0	0	0	0
AQ29	44	44	44	44	44
AQ30	44	44	44	44	44
Functionalities					
AQ31	-20	-8	-20	11	20
AQ32	22	18	18	12	0
AQ33	-20	-16	-20	11	-20
AQ34	22	-22	18	12	-22
AQ35	-20	-8	-20	-6	-20
AQ36	-20	-20	-20	11	-20
AQ37	-20	-20	-20	-6	-20
AQ38	-20	-20	-20	0	-20
AQ39	6	18	18	0	-22
AQ40	-12	-12	-12	4	20
AQ41	-25	25	25	0	0
AQ42	-22	-22	-22	-22	-22
AQ43	24	24	-24	24	-24
AQ44	24	24	-24	24	-24
AQ45	-24	-24	-24	-24	-24
AQ46	-16	-16	27	27	27
AQ47	-13	-13	-13	22	-13
AQ48	-13	-13	-13	22	22
Architecture and application					
AQ49	35	35	35	35	35
AQ50	29	29	29	29	29
AQ51	38	38	38	38	38
AQ52	29	29	29	29	29
AQ53	18	18	18	18	18
AQ54	29	29	29	29	29
Settings					
AQ55	27	27	27	27	-27
AQ56	13	13	13	13	13
AQ57	-15	-15	-15	-15	-15
Privacy policies					
AQ58	22	22	22	22	22
AQ59	-32	32	32	-32	-32
AQ60	-22	22	-22	22	-22
AQ61	32	32	32	-32	32

TABLE 18. *(Continued.)* Specific ScoreAQ for the investigated systems.

AQs	Facebook	Twitter	Tumblr	Diaspora	Gab
Feedbacks					
AQ62	-20	20	-20	-20	-20
AQ63	30	30	30	30	30
AQ64	30	30	30	30	30
API and third-party relationships					
AQ65	41	41	41	41	41
AQ66	-32	-32	-32	-32	-32
AQ67	-75	-75	75	75	-75
AQ68	-47	-47	-47	47	-47

45) Can the user search other users by interest?

46) Does the SUI detect the user's location?

47) Does the SUI recommend potential friends?

48) Does the SUI recommend potential interests?

- **Architecture and application**

49) Does the SUI keep up to date all the appliances/ Software/ Hardware on a frequent basis?

50) Does the SUI perform regular vulnerability assessments?

51) Does the SUI use layered protections?

52) Does the SUI provider use the security best practices in developing their applications?

53) Does the SUI developer update and patch regularly the application?

54) Does the SUI developer review and test the application for security vulnerabilities?

- **Settings**

55) Does the SUI allow the users to adjust their privacy settings?

56) Does the SUI provide automated privacy settings that learn from the user's behavior and preferences?

57) Are the settings clear and easy to use?

- **Privacy policies**

58) Does the SUI provide privacy policies?

59) Is yes, does the SUI inform the users about changes in the privacy policies?

60) Are the privacy policies written in clear and understandable terms?

61) Does the SUI update the policies regularly?

- **Feedback**

62) Does the SUI implement user's feedback about the usability of privacy settings?

63) Does the SUI allow the users to report malicious behavior?

64) Does the SUI allow the users to report compromised accounts or malicious users?

- **API and third-party relationships**

65) Does the SUI developer use trusted external libraries?

66) Does the SUI provide an API for developers to integrate the OSN services into their apps?

67) Does the SUI allow third-party applications to access and collect data?

68) Does the SUI allow users to ban a third-party application from accessing their information?

APPENDIX B THE INVESTIGATED SYSTEMS SCORES

Tables 17 and 18 compare the score obtained for each assessment question between the investigated systems.

REFERENCES

[1] Statista. (2019). *Number of Social Media Users Worldwide from 2010 to 2021 (in Billions)*. Accessed: Mar. 20, 2019. [Online]. Available: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[2] H. H. Khondker, "Role of the new media in the Arab spring," *Globalizations*, vol. 8, no. 5, pp. 675–679, 2011.

[3] P. Panagiotopoulos, A. Z. Bigdeli, and S. Sams, "Citizen government collaboration on social media: The case of twitter in the 2011 riots in England," *Government Inf. Quart.*, vol. 31, no. 3, pp. 349–357, Jul. 2014.

[4] Pleaserobme. (2019). *Raising Awareness About Over-Sharing*, accessed: 22-Mar-2019. [Online]. Available: http://pleaserobme.com/

[5] M. Bartsch and T. Dienlin, "Control your facebook: An analysis of online privacy literacy," *Comput. Hum. Behav.*, vol. 56, pp. 147–154, Mar. 2016.

[6] K. Lewis, J. Kaufman, and N. Christakis, "The taste for privacy: An analysis of college student privacy settings in an online social network," *J. Comput.-Mediated Commun.*, vol. 14, no. 1, pp. 79–100, Oct. 2008.

[7] IDP. *Gpen Privacy Sweep 2017 Finds Ambiguity in Privacy Policies*. Accessed: Sep. 6, 2018. [Online]. Available: http://www.idp.al/2017/10/25/gpen-privacy-sweep-2017-finds-ambiguity-in-privacy-policies/

[8] T. Stern and N. Kumar, "Improving privacy settings control in online social networks with a wheel interface, improving privacy settings control in online social networks with a wheel interface," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 3, pp. 524–538, Mar. 2014. [Online]. Available: http://doi.wiley.com/10.1002/asi.22994

[9] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2019–2036, Apr. 2014.

[10] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security issues in online social networks," *IEEE Internet Comput.*, vol. 15, no. 4, pp. 56–63, Jul./Aug. 2011.

[11] J. P. Albrecht. *Legislative Train Schedule | European Parliament*, Accessed: Feb. 17, 2019. [Online]. Available: http://www.europarl.europa.eu/legislative-train/theme-area-of-justice-and-fundamental-rights/file-general-data-protection-regulation

[12] *Home page of eu GDPR*, Accessed: Feb. 17, 2019. [Online]. Available: https://www.eugdpr.org/

[13] D. Hallinan, M. Friedewald, and P. Mccarthy, "Citizens perceptions of data protection and privacy in europe," *Comput. Law Secur. Rev.*, vol. 28, no. 3, pp. 263–272, 2012.

[14] T. Robertson. *Top Five Concerns With GDPR Compliance*. Accessed: Feb. 17, 2019. [Online]. Available: https://blogs.thomsonreuters.com/financial-risk/risk-management-and-compliance/top-five-concerns-gdpr-compliance/

[15] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *J. Comput.-Mediated Commun.*, vol. 13, no. 1, pp. 210–230, Oct. 2007.

[16] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," in *Proc. 9th WebKDD 1st SNA-KDD Workshop Web Mining Social Netw. Anal.*, Apr. 2007, pp. 56–65.

[17] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explorations Newslett.*, vol. 10, no. 2, pp. 12–22, Dec. 2008.

[18] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[19] P. Kumari, "Requirements analysis for privacy in social networks," in *Proc. 8th Intl. Workshop Tech., Econ. Legal Aspects Bus. Models Virtual Goods (VG)*, 2010, pp. 1–7.

[20] J. Buchmann, *Internet Privacy: Options for Adequate Realisation*. New York, NY, USA: Springer, 2013.

[21] B. Schneier, "A taxonomy of social networking data," *IEEE Secur. Privacy*, vol. 8, no. 4, p. 88, Aug. 2010.

[22] S. D. Warren and L. D. Brandeis, "The right to privacy," *Harvard Law Rev.*, vol. 4, no. 5, pp. 193–220, Dec. 1890.

[23] S. Patil and A. Kobsa, "Privacy considerations in awareness systems: Designing with privacy in mind," in *Human-Computer Interaction Series*. New York, NY, USA: Springer, 2009, pp. 187–206.

[24] *Privacy | Definition of Privacy in English by Oxford Dictionaries*. Accessed: Jan. 29, 2019. [Online]. Available: https://en.oxforddictionaries.com/definition/privacy

[25] L. Staff. *Privacy*. Accessed: Jan. 29, 2019. [Online]. Available: https://www.law.cornell.edu/wex/privacy

[26] Q. Ni, E. Bertino, J. Lobo, C. Brodie, C.-M. Karat, J. Karat, and A. Trombeta, "Privacy-aware role-based access control," *ACM Trans. Inf. Syst. Secur.*, vol. 13, no. 3, p. 24, 2010.

[27] J. Rachels, "Why privacy is important," *Philosophy Public Affairs*, vol. 4, no. 4, pp. 323–333, 1975. [Online]. Available: http://www.jstor.org/stable/2265077

[28] C. Bunnig and C. H. Cap, "Ad hoc privacy management in ubiquitous computing environments," in *Proc. 2nd Int. Conf. Adv. Hum.-Oriented Personalized Mech., Technol., Services*, Sep. 2009, pp. 85–90.

[29] S. Taheri, S. Hartung, and D. Hogrefe, "Achieving receiver location privacy in mobile ad hoc networks," in *Proc. IEEE 2nd Int. Conf. Social Comput.*, Aug. 2010, pp. 800–807.

[30] F. I. Processing, P. J. Bond, U. S. F. Technology, and A. L. Bement. *Standards for Security Categorization of Federal Information and Information Systems*, Accessed: Mar. 20, 2019. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.199.pdf

[31] M. Bishop, *Introduction to Computer Security*, 1st ed, 1st ed. Boston, MA, USA: Addison-Wesley, 2008.

[32] M. Hafner and R. Breu, "Basic concepts of SOA security," in *Security Engineering for Service-Oriented Architectures*. Berlin, Germany: Springer, 2009, pp. 27–45.

[33] S. Harris, *CISSP All-in-One Exam Guide*. New York, NY, USA: McGraw-Hill, 2012.

[34] *National Institute of Standards and Technology*. Accessed: Jan. 29, 2019. [Online]. Available: https://www.nist.gov/

[35] S. Brooks, E. Nadeau, M. Garcia, N. Lefkovitz, and S. Lightman, "Privacy risk management for federal information systems," NIST, Gaithersburg, MD, USA, Tech. Rep. NISTIR 8062, 2015.

[36] A. Cavoukian, "Privacy by design: The 7 foundational principles," Inf. Privacy Commissioner Ontario, Toronto, ON, Canada, 2009.

[37] I. Wagner and E. Boiten, "Privacy risk assessment: From art to science, by metrics," in *Cryptocurrencies Blockchain Technology*. New York, NY, USA: Springer, 2018, pp. 225–241.

[38] N. Vollmer. (2018). *Article 35 Eu General Data Protection Regulation (EU-GDPR)*. [Online]. Available: http://www.privacy-regulation.eu/en/article-35-data-protection-impact-assessment-GDPR.htm

[39] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, "Privacy-as-a-service: Models, algorithms, and results on the facebook platform," in *Proc. Web*, 2009, p. 2.

[40] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 1, p. 6, Dec. 2010.

[41] F. B. Baker and S.-H. Kim, *Item Response Theory: Parameter Estimation Technology*. Boca Raton, FL, USA: CRC Press, 2004.

[42] A. Srivastava and G. Geethakumari, "Measuring privacy leaks in online social networks," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2013, pp. 2095–2100.

[43] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris, "Pscore: A framework for enhancing privacy awareness in online social networks," in *Proc. 10th Int. Conf. Availability, Rel. Secur. (ARES)*, Jul. 2015, pp. 592–600.

[44] R. G. Pensa and G. D. Blasi, "A privacy self-assessment framework for online social networks," *Expert Syst. Appl.*, vol. 86, pp. 18–31, Nov. 2017.

[45] J. L. Becker and H. Chen, *Measuring Privacy Risk in Online Social Networks*. Oakland, CA, USA: Univ. California, 2009.

[46] T. H. Ngoc, I. Echizen, K. Komei, and H. Yoshiura, "New approach to quantification of privacy on social network sites," in *Proc. 24th IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Apr. 2010, pp. 556–564.

[47] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, "Privometer: Privacy protection in social networks," in *Proc. IEEE 26th Int. Conf. Data Eng. Workshops (ICDEW)*, Apr. 2010, pp. 266–269.

[48] C. Akcora, B. Carminati, and E. Ferrari, "Privacy in social networks: How risky is your social graph?" in *Proc. IEEE 28th Int. Conf. Data Eng.*, Apr. 2012, pp. 9–19.

[49] B. Vidyalakshmi, R. K. Wong, and C.-H. Chi, "Privacy scoring of social network users as a service," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jul. 2015, pp. 218–225.

[50] R. K. Nepali and Y. Wang, "Sonet: A social network model for privacy monitoring and ranking," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst. Workshops (ICDCSW)*, Jul. 2013, pp. 162–166.

[51] Y. Wang and R. K. Nepali, "Privacy measurement for social network actor model," in *Proc. Int. Conf. Social Comput.*, Sep. 2013, pp. 659–664.

[52] Y. Wang, R. K. Nepali, and J. Nikolai, "Social network privacy measurement and simulation," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 802–806.

[53] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, Feb. 2010, pp. 251–260.

[54] N. Z. Gong and B. Liu, "Attribute inference attacks in online social networks," *ACM Trans. Privacy Secur.*, vol. 21, no. 1, pp. 3:1–3:30, Aug. 2018.

[55] (2019). *PDCA*. [Online]. Available: https://en.wikipedia.org/wiki/PDCA

[56] V. R. Basili, "Software modeling and measurement: The goal/question/metric paradigm," Univ. Maryland, College Park, MD, USA, Tech. Rep. UMIACS TR-92-96, 1992.

[57] V. R. Basili, "Applying the goal/question/metric paradigm in the experience factory," *Softw. Qual. Assurance Meas. A, Worldwide Perspective*, vol. 7, no. 4, pp. 21–44, Oct. 1993.

[58] B. Sanz, C. Laorden, G. Alvarez, and P. G. Bringas, "A threat model approach to attacks and countermeasures in on-line social networks," in *Proc. 11th Reunion Espanola Criptografia y Seguridad Información (RECSI)*, Aug. 2010, pp. 343–348.

[59] Y. Wang and R. K. Nepali, "Privacy threat modeling framework for online social networks," in *Proc. Int. Conf. Collaboration technol. Syst. (CTS)*, Jun. 2015, pp. 358–363.

[60] G. NaliniPriya and M. Asswini, "A survey on vulnerable attacks in online social networks," in *Proc. Int. Conf. Innov. Inf. Comput. Technol.*, Feb. 2015, pp. 1–6.

[61] I. Kayes and A. Iamnitchi, "Privacy and security in online social networks: A survey," *Online Social Netw. Media*, vol. 3, pp. 1–21, Oct. 2017.

[62] S. Oukemeni, H. R.-P. Helena, and J. M. M. Puig, "Privacy analysis on microblogging online social networks: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 60:1–60:36, Jan. 2019. doi: 10.1145/3321481.

[63] P. Kumari, A. Pretschner, J. Peschla, and J.-M. Kuhn, "Distributed data usage control for Web applications," *Proc. 1st ACM Conf. Data Appl. Secur. Privacy*, Feb. 2011, pp. 85–96.

[64] E. Kozhemyak, "Privacy considerations for secure identification in social wireless networks," M.S. thesis, School Comput. Sci. Commun., Royal Inst. Technol., Stockholm, Sweden, 2011. [Online]. Available: https://bit.ly/2OmthaL

[65] M. B. Islam, "Privacy by design for social networks," Ph.D. dissertation, School Inf. Syst. Sci. Eng. Fac., Queensland Univ. Technol., Brisbane, QLD, Australia, 2014.

[66] J. Lalchandani and H. B. Sankaranarayanan, "Risk weighted social trust index for online social networks," *Procedia Comput. Sci.*, vol. 78, pp. 307–313, Sep. 2016. doi: 10.1016/j.procs.2016.02.060.

[67] Y. Farashazillah, "A security framework to protect data in cloud storage," Ph.D. dissertation, Univ. Southampton, Southampton, U.K., 2017. [Online]. Available: https://eprints.soton.ac.uk/415861/1/Final_Thesis.pdf

[68] *Open Web Application Security Project: OWASP Risk Rating Methodology*. Accessed: Jun. 25, 2019. [Online]. Available: https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology

[69] OWASP. *Owasp top 10 Privacy Risks Project*. Accessed: Jun. 25, 2019. [Online]. Available: https://www.owasp.org/index.php/OWASP_Top_10_Privacy_Risks_Project

[70] NIST. (2012). *Sp 800-30r1. Guide For Conducting Risk Assessments*. Accessed: Jun. 25, 2019. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf

[71] *Sp 800-53r4. Security and Privacy Controls for Federal Information Systems and Organizations*. Accessed: Jun. 25, 2019. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf

[72] *Common Vulnerabilities and Exposures (CVE)*. Accessed: Jun. 25, 2019. [Online]. Available: https://cve.mitre.org/

[73] E. Aghasian, S. Garg, L. Gao, S. Yu, and J. Montgomery, "Scoring users' privacy disclosure across multiple online social networks," *IEEE Access*, vol. 5, pp. 13118–13130, 2017.

[74] *Facebook*, Accessed: Jun. 20, 2019. [Online]. Available: https://www.facebook.com/

[75] *Twitter*. Accessed: Mar. 20, 2019. [Online]. Available: https://twitter.com/

[76] *tumblr*. Accessed: Jun. 20, 2019. [Online]. Available: https://www.tumblr.com/login

[77] *JoinDiaspora**.Accessed: Mar. 20, 2019. [Online]. Available: https://www.joindiaspora.com/

[78] *Gab*. Accessed: Jun. 20, 2019. [Online]. Available: http://gab.ai

[79] D. W. Stout. *Social Media Statistics: Top Social Networks by Popularity*. Accessed: Jun. 25, 2019. [Online]. Available: https://dustinstout.com/social-media-statistics/

[80] Y. Matsunobu. (2014). *Semi-Synchronous Replication at Facebook*. Accessed: Jun. 25, 2019. [Online]. Available: http://yoshinorimatsunobu.blogspot.com.es/2014/04/semi-synchronous-replication-at-facebook.html

[81] D. C. Knowledge. (2016). *The Facebook Data Center FAQ*. Accessed: Jun. 25, 2019. [Online]. Available: http://www.datacenterknowledge.com/the-facebook-data-center-faq/

[82] A. Rosen. (2017). *Tweeting Made Easier*. Accessed: Mar. 20, 2019. [Online]. Available: https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.html

[83] P. Schuller. (2014). *Manhattan, Our Real-Time, Multi-Tenant Distributed Database for Twitter Scale*. Accessed: Mar. 20, 2019. [Online]. Available: https://blog.twitter.com/2014/manhattan-our-real-time-multi-tenant-distributed-database-for-twitter-scale

[84] Twitter. (2018). *Privacy Policy*. Accessed: Mar. 20, 2019. [Online]. Available: https://twitter.com/en/privacy

[85] (2019). *Request to Verify an Account | Twitter Help Center*. Accessed: Mar. 20, 2019. [Online]. Available: https://support.twitter.com/articles/20174631

[86] Tumblr. (2018). *European Privacy Policy*. Accessed: Mar. 20, 2019. [Online]. Available: https://www.tumblr.com/privacy/en_eu

[87] Statista. (2019). *Cumulative Total of TUMBLR BLOGS from May 2011 to April 2019 (in Millions)*. Accessed: Jun. 20, 2019. [Online]. Available: https://www.statista.com/statistics/256235/total-cumulative-number-of-tumblr-blogs/

[88] T. Hoff. (2012). *Tumblr Architecture—15 Billion Page Views a Month and Harder to Scale than Twitter*. Accessed: Jun. 20, 2019. [Online]. Available: http://highscalability.com/blog/2012/2/13/tumblr-architecture-15-billion-page-views-a-month-and-harder.html

[89] MariaDB. (2018). *Tumblr Uses Mariadb for Multi-Source Replication*. Accessed: Jun. 20, 2019. [Online]. Available: https://mariadb.com/kb/en/mariadb/tumblr-uses-mariadb-for-multi-source-replication/

[90] Tumblr. (2018). *Tumblr Privacy Policy*. Accessed: Jun. 20, 2019. [Online]. Available: https://www.tumblr.com/policy/en/privacy

[91] *DiasporaFoundation*. Accessed: Mar. 20, 2019. [Online]. Available: https://diasporafoundation.org/

[92] G. Morehouse. (2015). *'Why Won't My Friends #joinDiaspora?' My Thoughts*. Accessed: Mar. 20, 2019. [Online]. Available: https://joindiaspora.com/posts/2fdcda606b2c0133cd8b2adb80a2c223

[93] Wikipedia. (2019). *Diaspora (Social Network)*. Accessed: Mar. 20, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Diaspora_(social_network)

[94] (2014). *JoinDiaspora: The Online Social World Where You are in Control*. Accessed: Mar. 20, 2019. [Online]. Available: http://www.joindiaspora.org/

[95] A. Dreyfus. (2017). *Gab.Ai Bans First User*. Accessed: Jun. 25, 2019. [Online]. Available: http://www.1776again.com/2017/01/26/gab-ai-bans-first-user/

[96] *GAP Pro*. Accessed: Jun. 25, 2019. [Online]. Available: https://gab.ai/pro

[97] A. Torba. *Andrew Torba on Gab*. Accessed: Jun. 25, 2019. [Online]. Available: https://gab.ai/a/posts/12463989

[98] *Privacy Impact Assessment (PIA)*. Accessed: Jun. 20, 2019. [Online]. Available: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf

SAMIA OUKEMENI received the Engineering degree in telecommunications and networks from the Institut National des Postes et Télécommunication, Rabat, Morocco, in 2010, and the M.S. degree in information systems security from Al Akhawayn University, Ifrane, Morocco, in 2016. She is currently pursuing the Ph.D. degree in networks and information technologies with the Universitat Oberta de Catalunya, under the supervision of Dr. H. Rifà-Pous and Dr. J. M. Marquès Puig. She is a Ph.D. grant holder from the same university.

After her engineering degree, she was an Information Security Engineer and the Project Manager with the Ministry of Interior, Rabat, until 2014. She was a part-time Leadership Program Assistant with the Leadership Development Institute (LDI), Al Akhawayn University.

HELENA RIFÀ-POUS received the Engineering degree in telecommunications, in 2001, and the Ph.D. degree in telematics engineering from the Universitat Politècnica de Catalunya (UPC), in 2008. She is currently an Associate Professor with the Faculty of Computer Science, Multimedia and Telecommunications, Universitat Oberta de Catalunya (UOC). She is also a member of the K-riptography and Information Security for Open Networks (KISON) Research Group. Her research interests include security and privacy protocols, with a special interest in distributed and wireless networks (e.g., in the context of smart cities).

JOAN MANUEL MARQUÈS PUIG graduated from FIB, Universitat Politècnica de Catalunya (UPC), Barcelona, in 1991, and received the Ph.D. degree from UPC, under the supervision of Prof. L. Navarro, in 2003.

From 1995 to 2013, he was a part-time Associate Professor with the Computer Architecture Department, UPC. Since 1997, he has been an Associate Professor of computer science studies with the Universitat Oberta de Catalunya (UOC). He is also a member of the Internet Computing and Systems Optimization (ICSO) Research Group, Universitat Oberta de Catalunya. His research interests include Internet scale systems, collaborative systems, scalable Internet services, cloud and peer-to-peer systems, privacy, collaborative learning, and networking.

Dr. Marquès Puig is a member of the Association for Computing Machinery (ACM).