Online Social Networks and Media 6 (2018) 18–25

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

Decentralized privacy preserving services for Online Social Networks

Leila Bahri[a], Barbara Carminati[b,*], Elena Ferrari[b]

[a] Royal Institute of Technology, Stockholm, Sweden
[b] Insubria University, Varese, Italy

ARTICLE INFO

Article history:
Received 3 October 2017
Revised 1 February 2018
Accepted 12 February 2018

Keywords:
Privacy preservation
Online Social Networks
Decentralized Online Social Networks
Privacy services

ABSTRACT

Current popular and widely adopted Online Social Networks (OSNs) all follow a logically centered architecture, by which one single entity owns unprecedented collections of personal data in terms of amount, variety, geographical span, and richness in detail. This is clearly constituting one of the major threats to users privacy and to their right to be-left-alone. Decentralization has then been considered as the panacea to privacy issues, especially in the realms of OSNs. However, with a more thoughtful consideration of the issue, it could be argued that decentralization, if not designed and implemented carefully and properly, can have more serious implications on users privacy rather than bringing radical solutions. Moreover, research on Decentralized Online Social Networks (DOSNs) has shown that there are more challenges to their realization that need proper attention and more innovative technical solutions. In this paper, we discuss the issues related to privacy preservation between centralization and decentralization, and we provide a review of available research work on decentralized privacy preserving services for social networks.

1. Introduction

Popular Online Social Networks (OSNs), such as Facebook, Twitter, or LinkedIn, are logically centralized services, that are owned and managed by single business entities. It is no secret to any one that their business, although presented as a *free-service* provision model, is fueled by targeted and retargeted marketing [2]. These are two marketing strategies that stand on data collection and on learning as much as possible about potential customers, their tastes, their habits, their spending patterns, or even their feelings and states of mind. Based on such information, potential customers can be smartly targeted, or retargeted by presenting the right product to the right person at ideally the right time, making advertisement more profitable. Most of these popular OSN companies count over hundreds of millions of registered users that enjoy their services, and occupy considerable portions of their computing and storage resources without direct subscriptions or monetary service fees. Therefore, they have a vital interest in collecting as much data as possible about their "free" subscribers, and in learning all what could possibly be extracted from this collected data. This collected data does not only cover the information that OSN users willingly upload and share with their contacts, but encompasses it to implicitly disclosed information, such as the times when users are online, the locations from where they connect, the type of activities they perform based on different locations and times, etc.

This unprecedented massive and uncontrolled collection and aggregation of different types of data about millions of individuals, from across all areas of the globe, in the hands of a few centralized entities is considered as one of the most serious and fundamental threats to users right to-be-left-alone; that is to their right to privacy.[1] This has been even more accentuated with the detected incidents of data leakage, either accidentally, due to attacks exploiting security breaches in those systems,[2] or intentionally to interested third parties such as secret services, or other interested companies. This makes some advocates spell out the fact that OSN users are not customers but rather the primary commercialized product under the business model of current centralized OSN providers.[3]

One of the most systematic and straightforward responses to mitigate this privacy dilemma of logically centralized services is to move to decentralized architectures. This idea has given birth to research under the area of what has been known as Decentralized

* Corresponding author.
E-mail address: barbara.carminati@uninsubria.it (B. Carminati).

https://doi.org/10.1016/j.osnem.2018.02.001
2468-6964/

[1] https://www.eff.org/issues/social-networks, https://www.privacyinternational.org/node/8

[2] https://yahoo.tumblr.com/post/150781911849/an-important-message-about-yahoo-user-security

[3] An example is this article on Facebook selling users data: http://www.telegraph.co.uk/technology/facebook/8917836/Facebook-faces-EU-curbs-on-selling-users-interests-to-advertisers

Online Social Networks (DOSNs), where the fundamental motivation is to mitigate privacy issues that are inherrent to the centralized model by desiging solutions that can provide similar online socializing functionality without the need of any one single *central* trusted entity [3]. Achieving this has been considered under two main conceptions. The first one consists at an architecture of multiple independent federated servers that provide the same OSN functionality, from which users can freely choose which to join and whom to trust, and between which users can freely and seamlessly switch without losing any of their advantages or functionality (e.g., [5–7]). The second conception takes decentralization to its extremes and consists at building peer-to-peer (P2P) networks of end users devices, with direct one-to-one interactions between them (e.g., [8–10]).

Decentralization, if achieved properly, brings the promise to inherently clear away the major privacy concerns related to the centralized model. However, almost all research efforts on DOSNs have shown that building social networking functionality under a decentralized architecture opens up more technical challenges than what it theoretically promises to solve. In addition to technical challenges related to developing functional services, such as searches and recommendations, instant messaging, and instant information sharing, that show similar performance and sophistication levels achieved in the centralized model, DOSNs do not come free of privacy concerns either [3]. Indeed, while decentralization solves the single aggregation and collection data point and the privacy concerns related to it, it also removes all the other protection mechanisms that were under responsibility of the blown central provider. Tasks such as content storage, access control management, data retrieval, data backup, failure management, and other data managerial tasks, that are under the sole responsibility of the central provider under the centralized model, become themselves decentralized, and thus distributed across all the peers in the decentralized system.

In this paper, we compare and discuss the privacy challenges related to decentralizing online social networks. We also provide a summarized but thorough review of the available literature on decentralized privacy preserving mechanisms for social networks. We discuss their strengths and shortcomings, and we highlight the areas where we think more research work is still needed.

The rest of this paper is organized as follows. In Section 2, we provide a discussion on privacy issues and how they interplay from centralized to decentralized social networks, and we formulate three main challenges to privacy in DOSNs. In Sections 3–5 we further explain each of the three identified challenges, we review related works under each of them, and we provide an analysis of their drawbacks and open research questions. Finally, we conclude the paper in Section 6.

2. Privacy from centralization to decentralization

Privacy is an elastic concept that can refer to different things based on the context where it is used, by whom, and for which purpose. One of the first systematic written discussions on the concept of privacy was made in 1890 by Samuel Warren and Louis Brandeis in an essay entitled The Right to Privacy [11]. The essay presented normative views on what the authors believed should be protected by the law under the scope of privacy. To the authors, privacy encompassed a spectrum that is larger than physical protection of one's home or one's physical property, and was defended as the right to be let alone [11]. Warren and Brandeis focused mostly on the press and on the publicity effects produced by the new emerging technological inventions of that time, such as photography and widely distributed newspapers. They shed the light on the possible invasion of a person's private life by vast public dissemination of personal information, referring to it as informational privacy. They argued that new technology made it essential to explicitly recognize a more general right to privacy that covered people's right to have control over how their thoughts, sentiments, and emotions could be shared with others. Recently, some new laws and regulations started recognizing information privacy as a formalized right that is under the protection of the law. For instance, the Personal Information Protection and Electronic Documents Act (PIPEDA)[4] in Canada, or the more recent European General Data Protection Regulation (GDPR) that will get in force starting from May 25, 2018.[5]

The relentless and rapid development of technological devices, networks, and hardware has not taken long before making the Internet available to almost all people all over the world, and not a privilege accessible only to some [12]. This may have been what shifted the online world to an era of connected individuals on a social level, after it was mostly a web of connected computers, systems, and corporations. What would be known as the social web has made it possible for people to connect, to produce, and to share information in ways that were not possible before [12]. It therefore comes at no surprise that OSNs have known the popularity and massive adoption that we see today.[6]

OSNs connect millions of people of all ages and backgrounds across all the globe and offer them an open space for autonomous exchange [12]. This experience of freedom of open personal and social expression, that spans over geographical borders, resulted in a deliberate sharing of information about oneself of which the consequences on one's privacy remain unfathomable and obscure to most of OSNs users. More importantly, OSN users are rarely aware of the amounts and types of data and meta-data being collected about them, or about the value of this data and the extent to which highly sensitive information could be extracted from analyzing it.

Being aware of the privacy threats related to the massive amounts of data collected about users by OSN providers, and with the related data leakages, either intended or unintended, that have been regularly observed, privacy advocates and researchers have called for an alternative design structure of OSNs. As such, the literature, as well as some business initiatives,[7] have seen a number of research efforts on the development of decentralized designs for OSNs, either using the federated or the P2P conceptions. However, it turned out that decentralization has also its complications and challenges regarding the management of privacy, although on a different level. We discuss this in what follows.

2.1. Online vs. offline privacy

In general, privacy management in OSNs could be discussed under two main perspectives, referred to as *online privacy* and *offline privacy* [13]. Online privacy refers to control over data sharing with a user's direct contacts, such as what information should be visible to whom. This is also technically known as *access control management*, or the mechanism by which users can organize their data and control its access, as well as how this is enforced and guaranteed by the system. As for offline privacy, it relates to ensuring privacy-aware control over one's thoughts, emotions, behavior, trends, and personality that could be extracted from analyzing the diverse and dense collection of data and meta-data that is generated by using the online socializing services.

Under the centralized model, online privacy is one of the services inherently offered by the OSN provider. Online privacy is the

[4] https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

[5] https://www.eugdpr.org/

[6] According to SmartInsights, the number of active users in Facebook alone exceeded 1.8 million users: http://www.smartinsights.com/social-media-marketing/

[7] Such as the Diaspora federated online social network.

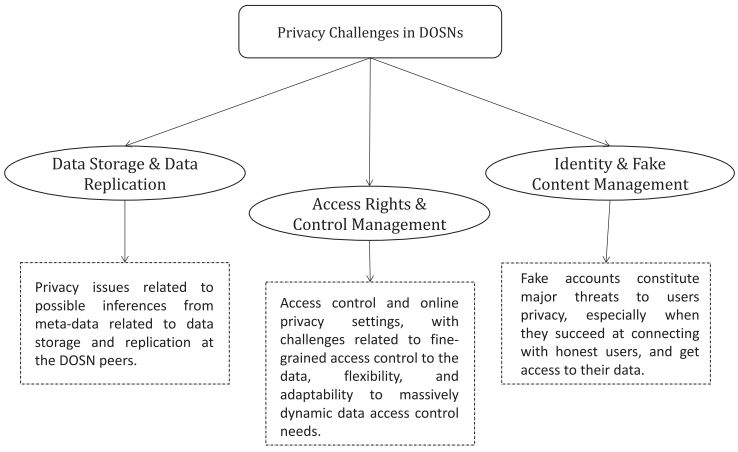

Fig. 1. Areas for privacy preservation services in DOSNs with related challenges.

closest level to users, and is the one that has more direct and more visible implications on them. For example, a user can see the direct implications and would be more concerned about having her night-party photo viewed by her office-mates, rather than the OSN provider profiling her as a depressed person and selling this information to an interested third party. For that, OSN providers have been paying considerable efforts at developing and enriching their offered interfaces for online privacy management. We have indeed seen substantial improvement in both the design, the functionality, and the granularity of privacy setting interfaces, especially in the most popular OSNs such as Facebook.[8]

Managing online privacy under the centralized model is technically more sound and easier, as all the data, and communication as well as access channels are centrally controlled by one owning entity. This also applies to federated architectures, as every federation (i.e., independent server) acts as a central node managing the data entrusted to it, including its access control. However, when it comes to P2P decentralization, this functionality of online privacy management moves from a centralized responsibility offered as a service by the OSN provider (or the federated server provider), to a shared and distributed responsibility among all the peers in the network. In addition to that, offline privacy issues may themselves not be simply eradicated by mere means of decentralization, creating more challenges to the creation of privacy preserving DOSNs.

We note that our focus in this paper is on decentralized P2P social networks, and we elaborate more on the related challenges in what follows.

2.2. Privacy challenges for DOSNs

Decentralization is supposed to answer offline privacy issues, as data is no more collected and owned by one central logical entity. However, this data distribution among peers in the decentralized network creates a new threat model with complicated technical challenges, especially when it comes to online privacy management. That is, access control and rights management becomes the distributed responsibility of the different peers that store a user's data. This makes coordination, and consensus agreement to maintain a secure state of the system one of the new required tasks and challenges of decentralization. In addition to that, decentralization in itself might not be enough to eradicate offline privacy concerns, as the different peers can themselves act as minor central points of control that can still analyze and learn from the data they store or they observe in the network [14]. This is even more challenging as the exchange of rich meta-data between peers is compulsory for the provision and management of basic OSN functionality. That is, it remains at stake how much information can be inferred from this managerial meta-data itself. Another privacy related issue that faces DOSNs is the management and control of fake accounts and of fake content. Although this might be considered as a system security issue, it has both direct and indirect consequences on users privacy. Indeed, fake accounts, if not properly labeled and detected can succeed at establishing valid connections with honest users, getting as such access to their personal information. In a DOSN, with the absence of a central authority, fake identities and malicious peers have more freedom to operate and infect the network without risking to be detected or removed, making this another challenge to ensuring privacy in the network.

In Fig. 1 we visualize the main challenging areas for privacy preservation in DOSNs with their descriptions. In the following sections, we address each of these identified challenges, we review their most prominent related works, and we provide insights and discussions on still open challenges.

3. Data storage and data replication

In DOSNs, information is not only stored at one single logically central location that is managed by one single known and accountable entity, but is rather dispersed across different peers of the P2P network. This dispertion is required mostly for availability and recoverability purposes. That is, the data of a user Alice, for instance, is replicated across multiple peers to ensure its availability even when Alice is offline or in case a failure happens at the level of her device [3]. Cryptography is therefore commonly used to blind data, both at rest and in communication. We note that although data encryption is also often used for access control and access rights management, as we elaborate in Section 4, its usage at this level concerns achieving data security at storage and replication nodes. That is, even if a DOSN assumes a non-cryptographic solution for data access rights management, the discussion here is regarding the confidentiality of data at storage points.

3.1. Centrally controlled data at peers level

While cryptography can ensure the confidentiality of users data, the corresponding meta-data needed for the data management, such as replication, ownership, and access requests handling, can in itself represent considerable privacy issues [15]. For example, data may be stored in encypted format at a peer, but this peer will still be able to observe its access patterns from other legtimate requestors. The data collected from such observation may be linked

[8] See for example, http://mattmckeon.com/facebook-privacy/

with other external information to identify the data owner. Moreover, as it has been discussed in [16], the peers in a DOSN might themselves be viewed as minor points of central control, with possibly more background information about the other peers, hence more potential for linking sensitive information about users.

In fact, one of the proposed strategies for data storage and replication in DOSNs is the use of direct friends and closer social peers, such as done in [10], where every user is placed within a shell (called matryoshka) of trusted direct friends who also serve as data replication, storage, and management nodes. Such a decision is double edged. From one side, it could be considered privacy wiser as people usually share their data with their direct and closer friends in any case. That is, direct friends and closer peers could be considered trusted by the data owner, thus privacy would not be an issue. However, those closer friends are also supposed to have more background knowledge about the data owner, and would therefore be able to better link external information to even encrypted content that they store in their devices. For example, the size of the data, the times when it is generated, the number, frequency, and trends of access requests to it, and other related metadata can implicitly contain sensitive information.

3.2. Selective storage solutions

One of the possible solutions to this issue of centrally controlled data at peers' level is to selectively choose the peers that will store some given data. For instance, in [10] it is suggested to store data only at those peers that are initially approved to have access to it. The drawback of such solutions is on performance and availability, as granted access is usually not uniformly distributed in data amount among all friends of a given user. For instance, it may be often that Alice shares more photos and content with Bob, compared to what she shares with all her other friends. In addition to that, central users with higher numbers of friends will be inherently overloaded, whereas other peripheral peers, that might as well have higher resources, will be under-utilized.

3.3. Summary and discussion

To the best of our knowledge, there are no works that directly tackle this issue of solving privacy implications related to data storage and replication in DOSNs. Probably this is due to the fact that DOSNs have not seen considerable adoption yet, keeping such issues still hidden to practitioners. However, with the new advances and the new technologies related to decentralized systems, such as Blockchains,[9] and the promises they bring to reviving social decentralized computing, we believe this is one of the research areas that requires more attention and effort. Indeed, there are already some starting initiatives that consider designing DOSNs using Blockchains, such as the Steem[10] and Synereo[11] projects, as well as some academic research discussions, such as in [17] and in [18]. These initiatives represent the start for a new, rich, and promising research area, where challenges such as how a Blockchain could best be used to ensure data storage security in a DOSN need to be explored and studied.

4. Access rights and control management

In a DOSN, every user locally holds her/his data and is only aware of their direct friends. Moreover, data dissemination is carried out in a peer to peer manner, making as such every peer responsible of managing access to their own data or to the data of other peers replicated on their nodes. As we have discussed under Section 3, it is a common practice in decentralized architectures to have data replicated between peers to ensure higher availability and to allow for data recovery in the event of a node's failure.

Whilst most of the proposals for DOSNs would opt for encryption to ensure the security of the out-sourced data at the end of the other peers, other proposals might overpass the need for encryption by making replicas only at the level of peers who are allowed to have access to it [3]. However, under both scenarios, a mechanism for the enforcement of access control within the decentralized environment is required and almost all the solutions available in the literature deploy encryption based techniques. Only a small subset of DOSN proposals come without encryption and rely on trust in friends only. An example is the work presented in [19] where authors suggest that users select a circle of trusted friends whose nodes would be used for both data replication and access control management. Clearly, this model assumes blind trust in the chosen circle of friends. Such an assumption makes it easier to address the access control and data security problems in DOSNs; however, it bases on an idealistic view that might not be practical for all users in OSNs.

Overall, access rights and control management for DOSNs has mainly been addressed using encryption based techniques. There are a few works that have presented proposals with a different approaches, such as audit-based access control. In the following, we review both the major works on encryption-based access control as well as the other few works that have presented non-encryption dependent approaches.

4.1. Encryption based solutions

Deploying data encryption to manage access control means that anyone could retrieve the encrypted content but only those who have the corresponding keys can interpret it. This implies that one of the requirements is to offer a mechanism for the distribution and management of the corresponding keys. Within the context of OSNs, characterized by massive amounts of data, huge users base, dynamic and frequently changing access requirements, etc., encryption keys management becomes an essential challenge [3]. Moreover, access scenarios in OSNs require fine grain levels of specification and span to cover relationships between users, between resources, and between users and resources. In addition to that, users shall be able to restrict access to their data in a fine-grained manner. As such, the main technical challenges related to the implementation of access control using encryption in OSNs are related to supporting fine-grain flexible access policies, with encryption schemes that offer the possibility of encrypting different combinations of single attributes for both single users and for groups of users. Moreover, given the highly dynamic nature of data sharing in OSNs, the scheme needs also to provide an efficient corresponding mechanism for the management of the related keys, in terms of distribution and revocation [4].

Therefore, most research works under this area have focused on finding usable ways for the dissemination, management, and revocation of the security keys related to the deployed encryption [10,20–23], or on deploying cryptography techniques that can offer finer levels of granularity, such as attribute based cryptography [9,21,22].

4.1.1. Keys dissemination

Regarding the issue of keys distribution, most of the proposals in the literature assume that the exchange of all access keys (or fingerprints that could be used to retrieve the keys) between users happens outside of the system (i.e., out of band – OOB) [3]. This

[9] Blockchain is a distributed ledger technology that is used to reach consensus on system-state in a purely P2P network. Blockchains have first started with the crypto-currency Bitcoin and are currently attracting much research attention for applications in different areas.

[10] https://steem.io/

[11] http://www.synereo.com/

results in the inconvenience of finding an appropriate and trustworthy OOB channel through which this keys exchange could take place. In a trial to overcome this problem, some proposals suggested the consideration of trusted nodes or super nodes in the DOSN to act as credential authorities [8,10]. These super nodes are only used to initiate the DOSN and are not implicated in the mediation of communication between nodes; hence cannot trace interactions. However, this approach still assumes the trustworthiness of these super nodes and their availability. As an alternative, some other works have suggested the exploitation of dynamic hash tables (DHTs) to include the management and distribution of keys (e.g., the proposal in [24]). This cancels the reliance on any central node, but does not provide any kind of identification. That is, the DHT is solely used to distribute the keys and the generation of identities requires to be managed with a different parallel service. However, the real problem and still open challenge with all these proposals is with the efficient management of keys revocation and/or update that should accompany the change of users access policies, especially given that such changes are so frequent and unpredictable in the realms of OSNs.

4.1.2. Dynamic access-group management

This issue of keys distribution and revocation gets more challenging and more complicated as it needs to be aligned with the support of fine-grain and flexible access policies, that are also highly changing. For instance, one of the main issues of decentralized access control in OSNs using encryption is groups membership management. Indeed, OSN users are often sharing common content with groups of friends that are highly dynamic. For instance, Bob might start sharing a photo with a group that contains Alice and 10 other friends, and then decides to revoke access from Kate and grant it to Jane instead. Such a scenario is expected to result in complex revocations and redistribution of keys that may affect all the members of the access group, and that may not be efficient and timely enough to provide instantaneous access revocation from Kate. To address these problems, different types of encyrption have been investigated and combined, such as, public key encryption (PKE), attribute based encryption (ABE), threshold secret sharing schemes, etc. [3].

PKE, also known as asymmetric cryptography, uses two kinds of keys: public and private. Public keys may be disseminated widely, whilst private keys are known only to the owner. In PKE, any person can encrypt a message using the public key of the receiver. Such a message can be decrypted only with the receiver's private key. In contrast to the more basic symmetric encryption, that relies on the same key to perform both encryption and decryption, PKE is considered to be more costly in its encryption/decryption processes. As such, it is common to use PKE only at the initiation of a secure communication, during which a symmetric key is exchanged between the communicating parties and it is thereafter used for the remaining of the established secure exchange. In the context of providing access control in DOSNs, one of the direct solutions is to use PKE to allow selective communication with every user. That is, every user in the network would be known by a public key that is distributed to all her friends. The friends will encrypt all information to be shared with a specific user with her public key. A similar idea has been proposed in [25], where the notion of relationship attestations was suggested. That is, the system generates a certificate for every two users who established a relationship in the network. These certificates could then be used, across multiple platforms that support the suggested protocol, to prove the existence of the attested relationship. Using direct PKE applications to address the access control problem in DOSNs is clearly not scalable as information needs to be encrypted for each friend it is to be shared with. This also will not allow revocation of access, as the entity that controls the decrypting private key is not the information owner but the receiver.

In practice, PKE is used as the basis for devising more encryption types that support more fine grain operations, such as ABE. Indeed, ABE is a type of PKE in which the secret key and ciphertexts are generated based on some given attributes (e.g., user role, country of residence, subscription type, etc.). That is, instead of relying on encrypting data for a given decrypting key that is related to a fixed identity, an encryption is made by labeling the generated ciphertext with a set of pre-defined attributes. Therefore, the decryption of a ciphertext is possible only if the attributes of a user key match the attributes of the secure text. This inherently requires that all users own a profile for which the credentials should satisfy the ciphertext attribute conditions to be able to decrypt a the message.

For the flexibility and fine grain support that it provides, ABE has been the most encryption type explored for access control on DOSNs. For instance, in [21], the authors adopted ABE where every user in the DOSN generates an ABE public key (APK) and a master secret key (AMSK). The notion of groups is then used to map to the attributes used in the encryption. As such, a user u would create for every friend f an ABE secret key (ASK) that corresponds to the set of attributes that express the groups that u assigns to f. However, one of the issues here is with user revocation, since the access policies are defined only over user attributes.

In order to offer support for revocation and dynamic group membership management, proxy encryption techniques that take advantage of the selective attribute group key distribution on top of ABE have been proposed. For instance, in [20], the authors adopted ABE and worked on offering support for dynamic group memberships and management of access right revocation without the need of reissuing keys or re-encrypting the data. They achieved this by introducing a proxy that needs to be contacted in order to be able to execute any decryption in the system. Users send the target ciphertext to the proxy that runs a transformation process on it. The transformed ciphertext can only be decrypted if the access right has not been revoked from the requesting user. Similarly, the authors in [1], demonstrate a proxy based system architecture that supports dynamic group membership and user revocation using ABE based access control policies. The work in [1] also provides support for escrow secrecy; that is, the proxy is also not able to decipher the content. They achieve this by relying on secure two-party computation protocols between the proxy and the data storage entity for the generation of required keys.

4.1.3. Collectively owned content

Another issue in access control in OSNs is related to collaboratively owned content. For instance, a post message made by user Bob where user Alice is also mentioned. One of the works we could find that address this issue for DOSNs is in [23]. The authors worked on the problem of collaboratively owned content, and access control issues related to it. More specifically, they have based on the problem of group photos that should be collaboratively owned by all the people in the group and not only by the person who shares the photo in the social network. Therefore, all the photo stakeholders should have a say in setting privacy rules for access to the photo they show on. To address that, the authors used a threshold-based sharing scheme, which is a cryptographic technique for the management of sharing a secret among different parties. That is, the content to secure is encrypted and is distributed in chunks among the different stakeholders. Each stakeholder on its own cannot reconstruct the distributed content (i.e., the shared secret). It is only when a required threshold of stakeholders come together and collaborate to combine the chunks they own that the content could be reconstructed.

4.2. Non encryption based solutions

The premise behind encryption based access control is to securely lock all pieces of users data and to allow the dissemination of appropriate keys to those allowed to gain access to it. This line of actions may provide deterministic solutions to the challenge of remote access control enforcement at the level of other peers, but fails at its best to provide the needed flexibility in both the formulation of fine grain access policies, and in their corresponding dissemination [20,26].

Indeed, information sharing in social networks has been observed to be characterized by denseness, diversity, and dynamicity. People share huge amounts of pieces of information under diverse formats (e.g., text, photos, voice) in a stochastic manner with continuously changing needs w.r.t dissemination preferences [27]. Moreover, information ownership does not follow a direct one-to-one fashion. In contrast to single-ownership of data where one definite entity is known to hold complete control over it, data ownership in social networks is mostly collaborative and multiple. For instance, a person might share a photo that displays other people, or might share a text in which other people are tagged, or might comment on a video shared by another person. All these elements make the firmness and rigidness of encryption mechanisms insufficient, at best, in providing an acceptable balance between protection and usability and efficiency of an access control solution for DOSNs.

A few research works in the literature have tried to approach the problem from a non encryption based perspective, by relying on trust-based approaches. For instance in [10], the authors have suggested a data management architecture where each user is placed in the center of a shell made of her direct trusted friends. Trusted friends are thus the ones who provide data management and access controls. The drawback of this solution is with data availability. Usually, a user's friends will mostly be from the same geographical location and would be expected to be online and offline at almost the same times. This will make the user's data unavailable when the friends are offline. Moreover, this may also introduce privacy issues, as a user might trust some friends to see some of her data, but not to take care of managing access to it and observing her data's access patterns, such as previously discussed in Section 3.

Within the same approach of non encryption based solutions, we find the aposteriori audit-based access control solution proposed in [28]. In [28], the authors have devised a system based on an open sharing environment where all transactions are logged, and a third party trusted auditor performs post control to detect dishonest peers. The work adopts the scheme in [29] for modeling access rules based on relationship paths information related to the type of the relationship, the cumulative trust and the length of the considered path between resource owner and potential receiver. Resources are annotated with access rules and with an additional audit log that traces their sharing path and its descriptive values in terms of its length, the type of its relationships, and its cumulative trust. Data sharing is governed by means of both reporting of detected bad behavior, and by regular audits that take place by a trusted auditor. The system relies on a reputation management model by which honest behavior is encouraged and malicious nodes are segregated by means of low reputation scores.

4.3. Summary and discussion

To sum up, access control for DOSNs have mainly been addressed using encryption-based techniques. The available proposals under this appraoch demonstrate that there is room for devising encyrption mechanisms that may be used to solve the access control problem in DOSNs, with support of some level of flexibility and fine-grain policies. However, given the amounts of data and the number of connections users maintain in OSNs, and with the required granularity and changing requirements for privacy settings, encryption mechanisms, thus far, may not overcome their by-design inherent limitations, especially in terms of efficiency in managing instantaneous revocations and highly dynamic group memberships [8]. Some other research works have taken the approach of relying on trust-based or audit-based access control management. One of the promising research paths along this direction of trust-based and auditable access control is the investigation of Blochckains technology. Blockchains allow the achievement of secure consensus on, and tamper proof recording of system states in a puerly decentralized P2P environment [30]. As such, Blockchains could be used for both the recording of access control policies, as well as for the management of access requests with decentralized consensus on who is allowed to access what data. One of the early works providing an initial discussion on the topic is available in [31]. The paper provides an overview architecture for a possible usage scenario of Blockchains as means of consensus on data access rights, their management, and their auditing. The work remains an initial attempt that deserves further development and consideration.

5. Identity and fake content management

The detection of fake accounts and of fake content in an OSN is of crucial importance as far as both security and privacy of the users go. When the OSN environment contains undetected fake accounts, they might easily fool honest users into befriending them, making them as such able to access information that the honest users intend to share with their real friends only. The literature contains a plethora of research works on fake accounts in OSNs, mainly under the research area known as Sybil detection [32]. Most of the available techniques assume a centralized architecture where information about all users and connections can be parsed and analyzed. That is, fake accounts and fake content are mainly detected based on differentiating between their behavioral and structural trends and those of honest users [33]. Clearly, such a differentiation requires the application of learning techniques on central collections of data. This becomes quite challenging when it comes to decentralized settings where it is hard to track patterns and form representational models. There are a few works addressing fake accounts management in DOSNs. These can be mainly organized under two bodies of work. The first one relies on decentralized fake accounts detection, whereas the second takes the approach of validating identities in decentralized settings.

5.1. Decentralized fake accounts detection

Detecting fake accounts in decentralized P2P systems have been attracting much attention from different research communities, especially in mobile networks or in Internet-of-Things (IoT) (e.g., [34,35]). In these settings, peers are usually identified based on a formal identity management model, and honest users are marked by identity verification guarantees offered by the identity provider in the system. In DOSNs (as well as in OSNs), where identity is a loose concept and where users can join based on simple ownership of a valid email address, the task becomes more challenging.

The work in [36] presents a gossip-based model for behavioral group identification in a DOSN. While this work is not directly aimed at detecting fake accounts, the behavioral group identification mechanism could be leveraged to generate representations of honest behavioral patterns, that could aid in the detection of fake accounts.

Another work that provides a solution aligned with decentralized fake content detection is in [37]. The authors developed an

unsupervised and decentralized anti-spam detection mechanisms that is also resilient to malicious nodes participation. Using collaborative learning, the solution creates a validation overlay that assesses the credibility of information exchanged and excludes the misbehaving nodes from the system.

5.2. Decentralized identity validation

Identity validation in OSNs refers to the estimation of an account's trustworthiness. The goal is not to detect fake accounts, but is rather to label identities based on their profiles and on their behavior within the OSN communities or groups they would like to join and connect with. The goal is to create a social representational layer of common identity trends within identified groups or communities in the OSN, that users can use as directives for assessing the credibility of the potential accounts they can connect with. For instance, the work in [38] relies on gossip-based unsupervised learning to build representations of common identity trends within every detected community in the DOSN. The work relies on gossip-learning based on one-to-one exchange of locally learned models among the DOSN peers. Nodes collect information about their direct friends and learn the common identity trends, in terms of profile attributes and shared behavior patterns, among them. After rounds of exchange of these local models, the system converges to a state by which every node is aware of the different identity models present in the social communities it belongs to. As a result, every node has an identity representative model of honest users, based on which it can asses the trustworthiness of new nodes in its network of connections.

5.3. Summary and discussion

In general, this area of identity validation and defense mechanisms against fake accounts in DOSNs could be considered as under investigated. As we have discussed, the few works that could be related to decentralized fake accounts detection do not specifically target the problem as per se. This clearly points to the need for more elaborate studies and research works on mechanisms that could be exploited to fight against fake accounts in DOSNs. The same applies to identity validation as well where, to the best of our knowledge, only the work in [38] is available. This work in itself could be considered as only a start towards understanding identity trends and deploying them for identity validation in a pure P2P network. Therefore, more research attention is needed to study other possible mechanisms for identity validation, where peers in a DOSN collaborate to maintain the sanity of the environment, and minimize the chances for Sybil attacks, or allow their detection.

6. Conclusion

Decentralization is undoubtedly one of the inherent solutions to major offline privacy issues in OSNs; however, it does not come free of new challenges and issues to privacy itself. In this paper, we have presented a discussion on privacy issues related to the shift from centralized to decentralized architectures for social networks. We made our synthesis based on three main axes that we believe are of high importance when considering privacy preserving DOSNs. Namely, we have discussed the challenges related to data storage and replication, to data access control management, and to fake accounts and fake content management. We have provided a review of the major works that focus on providing privacy preserving services under decentralized conceptions of OSNs, and that operate under those identified axes. We have also highlighted areas where we believe there is still need for more research efforts, especially w.r.t dynamic group membership management and efficient and timely access revocation support. With the development of new technologies for P2P computing, such as Blockchains, that promise the provision of secure and strong platforms for accountable decentralized systems, we think that there is more potential around DOSNs research as well.

References

[1] H. Junbeom, Improving security and efficiency in attribute-based data sharing, Trans. Knowl. Data Eng. 25 (2013) 2271–2282.
[2] D.M. Scott, The New Rules of Marketing and PR: How to Use Social Media, Online Video, Mobile Applications, Blogs, News Releases, and Viral Marketing to Reach Buyers Directly, John Wiley & Sons, 2015.
[3] T. Paul, A. Famulari, T. Strufe, A survey on decentralized online social networks, Comput. Netw. Int. J. Comput. Telecommun. Netw. 75 (2014) 437–452.
[4] F. Günther, M. Manulis, T. Strufe, Cryptographic treatment of private user profiles, in: Proceedings of the International Conference on Financial Cryptography and Data Security, Springer, 2011, pp. 40–54.
[5] L. Schwittmann, C. Boelmann, M. Wander, T. Weis, Sonet–privacy and replication in federated online social networks, in: Proceedings of Distributed Computing Systems Workshops (ICDCSW), IEEE, 2013, pp. 51–57.
[6] A. Shakimov, H. Lim, R. Cáceres, L.P. Cox, K. Li, D. Liu, A. Varshavsky, Vis-a-vis: privacy-preserving online social networking via virtual individual servers, in: Proceedings of Third International Conference on Communication Systems and Networks (COMSNETS), 2011, IEEE, 2011, pp. 1–10.
[7] A. Bielenberg, L. Helm, A. Gentilucci, D. Stefanescu, H. Zhang, The growth of diaspora-a decentralized online social network in the wild, in: Proceedings of the 2012 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2012, pp. 13–18.
[8] S. Buchegger, D. Schiöberg, L.H. Vu, A. Datta, Peerson: P2p social networking: early experiences and insights, in: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems, ACM, 2009, pp. 46–52.
[9] S. Jahid, S. Nilizadeh, P. Mittal, N. Borisov, A. Kapadia, Decent: a decentralized architecture for enforcing privacy in online social networks, in: Proceedings of Pervasive Computing and Communications Workshops (PERCOM Workshops), IEEE, 2012, pp. 326–332.
[10] L.A. Cutillo, R. Molva, T. Strufe, Safebook: feasibility of transitive cooperation for privacy on a decentralized social network, in: Proceedings of World of Wireless, Mobile and Multimedia Networks & Workshops, 2009, WoWMoM'09, IEEE, 2009, pp. 1–6.
[11] S.D. Warren, L.D. Brandeis, The right to privacy, Harv. Law Rev. 4 (5) (1890) 193–220.
[12] N.B. Ellison, D. Boyd, Sociality through social network sites, in: The Oxford Handbook of Internet Studies, Oxford University Press, 2013, pp. 151–172.
[13] M. Bartsch, T. Dienlin, in: Control your facebook: an analysis of online privacy literacy, 56, Elsevier, 2016, pp. 147–154.
[14] S. Taheri-Boshrooyeh, A. Kupcu, O. Ozkasap, Security and privacy of distributed online social networks, in: Proceedings of Distributed Computing Systems Workshops (ICDCSW), IEEE, 2015, pp. 112–119.
[15] B. Greschbach, Privacy issues in decentralized online social networks and other decentralized systems, KTH Royal Institute of Technology, 2016 Ph.D. thesis.
[16] M. Qamar, M. Malik, S. Batool, S. Mehmood, A.W. Malik, A. Rahman, Centralized to decentralized social networks: factors that matter, in: Managing and Processing Big Data in Cloud Computing, IGI Global, 2016, pp. 37–54.
[17] D. Fu, L. Fang, Blockchain-based trusted computing in social network, in: Proceedings of 2016 2nd IEEE International Conference on Computer and Communications (ICCC), IEEE, 2016, pp. 19–22.
[18] F. Buccafurri, G. Lax, S. Nicolazzo, A. Nocera, Tweetchain: an alternative to blockchain for crowd-based applications, in: Proceedings of International Conference on Web Engineering, Springer, 2017, pp. 386–393.
[19] R. Narendula, T.G. Papaioannou, K. Aberer, A decentralized online social network with efficient user-driven replication, in: Proceedings of Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), IEEE, 2012, pp. 166–175.
[20] S. Jahid, P. Mittal, N. Borisov, Easier: encryption-based access control in social networks with efficient revocation, in: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ACM, 2011, pp. 411–415.
[21] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, D. Starin, Persona: an online social network with user-defined privacy, ACM SIGCOMM Comput. Commun. Rev. 39 (2009) 135–146.
[22] S. Nilizadeh, S. Jahid, P. Mittal, N. Borisov, A. Kapadia, Cachet: a decentralized architecture for privacy preserving social networking with caching, in: Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, ACM, 2012, pp. 337–348.
[23] P. Ilia, B. Carminati, E. Ferrari, P. Fragopoulou, S. Ioannidis, Sampac: socially-aware collaborative multi-party access control, in: Proceedings of the Seventh ACM Conference on Data and Application Security and Privacy, CODASPY '17, ACM, 2017, pp. 71–82. 10.1145/3029806.3029834
[24] K. Graffi, S. Podrajanski, P. Mukherjee, A. Kovacevic, R. Steinmetz, A distributed platform for multimedia communities, in: Proceedings of Tenth IEEE International Symposium on Multimedia, ISM'08, IEEE, 2008, pp. 208–213.
[25] A. Tootoonchian, S. Saroiu, Y. Ganjali, A. Wolman, Lockr: better privacy for social networks, in: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, ACM, 2009, pp. 169–180.

[26] A. Shakimov, A. Varshavsky, L.P. Cox, R. Cáceres, Privacy, cost, and availability tradeoffs in decentralized osns, in: Proceedings of the 2nd ACM Workshop on Online Social Networks, ACM, 2009, pp. 13–18.
[27] M. Netter, M. Riesner, M. Weber, G. Pernul, Privacy settings in online social networks–preferences, perception, and reality, in: Proceedings of 46th Hawaii International Conference on System Sciences (HICSS), IEEE, 2013, pp. 3219–3228.
[28] L. Bahri, B. Carminati, E. Ferrari, Cards-collaborative audit and report data sharing for a-posteriori access control in dosns, in: Proceedings of IEEE Conference on Collaboration and Internet Computing (CIC), IEEE, 2015, pp. 36–45.
[29] B. Carminati, E. Ferrari, A. Perego, Rule-based access control for social networks, in: Proceedings of the on the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Springer, 2006, pp. 1734–1744.
[30] F. Tschorsch, B. Scheuermann, Bitcoin and beyond: a technical survey on decentralized digital currencies, IEEE Commun. Surv. Tutor. 18 (3) (2016) 2084–2123.
[31] G. Zyskind, O. Nathan, et al., Decentralizing privacy: using blockchain to protect personal data, in: Proceedings of the 2015 IEEE Security and Privacy Workshops (SPW), IEEE, 2015, pp. 180–184.
[32] M. Al-Qurishi, M. Al-Rakhami, A. Alamri, M. Alrubaian, S.M.M. Rahman, M.S. Hossain, in: Sybil defense techniques in online social networks: a survey, 5, 2017, pp. 1200–1219.
[33] K. Anand, J. Kumar, K. Anand, Anomaly detection in online social network: a survey, in: Proceedings of International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE, 2017, pp. 456–459.
[34] M. Ehdaie, N. Alexiou, P. Papadimitratos, Random key pre-distribution techniques against sybil attacks, J. Commun. Eng. 5 (1) (2016) 1–13.
[35] P. Kavitha, C. Keerthana, V. Niroja, V. Vivekanandhan, Mobile-id based sybil attack detection on the mobile adhoc network, Int. J. Commun. Comput. Technol. 02 (02) (2014) 6–9.
[36] N. Laleh, B. Carminati, E. Ferrari, S. Girdzijauskas, Gossip-based behavioral group identification in decentralized osns, in: Machine Learning and Data Mining in Pattern Recognition, Springer, 2016, pp. 676–691.
[37] A. Soliman, S. Girdzijauskas, Dlsas: distributed large-scale anti-spam framework for decentralized online social networks, Proceedings of IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), IEEE, 2016, pp. 363–372.
[38] A. Soliman, L. Bahri, B. Carminati, E. Ferrari, S. Girdzijauskas, Diva: decentralized identity validation for social networks, in: Proceedings of 2015 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM), IEEE/ACM, 2015, pp. 383–391.

Leila Bahri has received a PhD in computer science from the University of Insubira, Italy where she has worked on designing alternative access control and identity management solutions for decentralised online social networks. She is currently a postdoc researcher at the Royal Institute of Technology (KTH) in Stockholm Sweden, where she works on a project for designing light-weight and energy efficient consensus algorithms for the Blockchain technology to make it more suitable for non-monetary applications, such as data provenance, data governance, identity management, etc. Her current research focus is on designing solutions for user-aware privacy management, and to explore technologies that could be in support for the implementation of privacy regulations and laws, such as the European GDPR.

Barbara Carminati is an associate professor of Computer Science at the University of Insubria, Italy, where she is the scientific director of the K& SM Research Center. She holds a PhD in Computer Science from the University of Milano, Italy. Her main research interests are related to security and privacy for innovative applications, like IoT, secure cloud computing, web services, data streams, and online social networks. Barbara Carminati has been involved in the organization of several international conferences as program committee member as well as program and general chair.

Elena Ferrari is a full professor of computer science at the University of Insubria, Italy, where she leads the STRICT Socialab. Her research activities are mainly related to data security, privacy and trust. Ferrari has a PhD in computer science from the University of Milano, Italy. She received the IEEE Computer Society's 2009 Technical achievement award for 'outstanding and innovative contributions to secure data management.' She is an IEEE fellow (for contributions to security and privacy for data and applications) and an ACM Distinguished Scientist.