Online Social Networks and Media 7 (2018) 12–29

Contents lists available at ScienceDirect

Online Social Networks and Media

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

Managing social contents in Decentralized Online Social Networks: A survey

Barbara Guidi[a,*], Marco Conti[b], Andrea Passarella[b], Laura Ricci[a]

[a] Department of Computer Science, University of Pisa, Largo Bruno Pontecorvo, Pisa, Italy
[b] Institute for Informatics and Telematics, IIT, CNR, Via Moruzzi, Pisa, Italy

ARTICLE INFO

Article history:
Received 13 February 2018
Revised 4 July 2018
Accepted 7 July 2018

Keywords:
Decentralized Online Social Networks
Data Management
Data availability
Information diffusion
Privacy

ABSTRACT

The widespread diffusion of Online Social Networks has given unforeseen opportunities for their users to share contents and mutually interact. However, current platforms offer inadequate guarantees as far as concerns the privacy of their users. To address these issues and leave to the users more control on their data, several recent proposals suggest to decentralize the storage of social data, aiming at leaving their control entirely to their owners. This approach has lead to the definition of Decentralized Online Social Networks (DOSNs), ranging from completely decentralized solutions, i.e. P2P solutions, to hybrid systems integrating external and private resources for storing user data. While DOSNs allow users to have more control over their data, they raise new challenges concerning the management and availability of social data. Guaranteeing data availability in a highly dynamic environment, defining proper algorithms for information diffusion and guaranteeing data privacy in a distributed setting are currently open problems in this area.

This survey presents an overview of these challenges and of the main solutions presented in the literature. Existing proposals are classified taking into account the strategies adopted to manage social data, focusing on the data availability and on information diffusion. The survey also presents the new privacy issues arising in DOSNs. Finally, an overview of the main open issues in this research area is presented.

1. Introduction

Online Social Networks (OSNs) are nowadays one of the most popular applications in the Internet. They have attracted a huge amount of users during the last years by changing the way people communicate and interact. Facebook can be considered the most representative OSN with, at the beginning of 2018, more than 2 billion active users, and the highest number of daily users connections. OSNs provide several services [1] offering to their users the opportunity of building a public profile, looking up new friends among the registered users, establishing relationships, and sharing content. Furthermore, these platforms also allow sharing of information within groups of users and the possibility of building communities of users characterized by common interests.

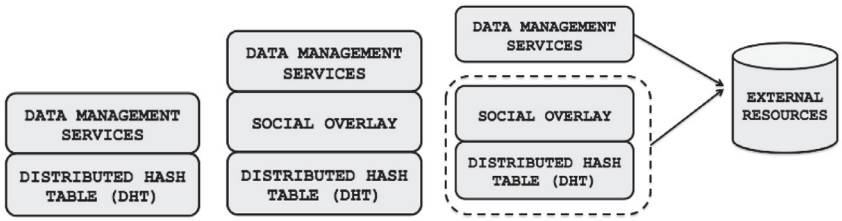
One of the major problems of current OSNs, which are mainly developed on centralized platforms, concerns the privacy of the users' data. Indeed, social data are stored in centralized servers, and the companies running the OSNs, use these data for commercial goals. In May 2015, a report[1] commissioned by the Belgian Data Protection Authority based on the analysis of OSN policies and terms-of-use, concluded that Facebook gives users a false sense of control over their data privacy. More recently, it has been evident to the general public that Facebook data might have been sold without any consent of the legitimate owners. These are only few examples of several legal issues that involve not only Facebook, but also further OSNs, like Twitter, or Google+. Furthermore, centralized OSNs may suffer of other problems, like limited scalability and high maintenance costs to manage data of large number of users [2].

All these issues have led researchers to propose alternative solutions based on the decentralization of OSN services. A Decentralized Online Social Network (DOSN) [3] is an online social network implemented on a distributed platform. In a DOSN, there is no single service provider but a set of nodes that cooperate to guarantee all the functionalities offered by a centralized OSN. Decentralization gives several benefits in terms of privacy. Indeed, there is no central entity that has the control on all users' data or changes the

* Corresponding author.
E-mail address: guidi@di.unipi.it (B. Guidi).

[1] https://www.huntonprivacyblog.com/tag/belgium/.

https://doi.org/10.1016/j.osnem.2018.07.001
2468-6964/



(a) DHT-based (b) SO-based (c) External Resources-based

Fig. 1. Distributed Online Social Network architectures.

existing terms of service, and this gives to the users more control over their data.

Moving from a centralized to a distributed architecture gives the opportunity to develop the social network platform by exploiting different distributed models. Following the way of how people interact with each other in a social network makes P2P architectures [4] a natural way to implement DOSNs. However, several alternative solutions are possible, like exploiting network of trusted servers, or mobile and opportunistic networks. Also hybrid solutions where the user exploits both the storage of its own device and a cloud storage service are possible [3].

Researchers have presented several solutions for DOSNs in the last decade. Furthermore, recent years have seen several initiatives to implement and deploy real DOSNs. Just to provide a few examples, the precursor of current DOSNs can be considered Diaspora [5], a solution based on a federation of trusted servers, which has gained a lot of popularity, despite not fully decentralized. Other DOSNs, like Tent and Friendica, are based on similar concepts. On the other hand, Retroshare is a fully decentralized DOSN, designed to provide maximum security and anonymity to its users.

While decentralization gives interesting possibilities for increasing the privacy level of users' data, it introduces many challenges, still to be solved. These mainly concern the management of social data in a distributed environment. With the term "social data", we identify all data exchanged in the Social Networks concerning both information related to the users (contact details, describing the userâs identity, relationships, community memberships, etc..) and generated contents (comments, posts, etc..). One of the main problems is to guarantee the availability of social data, in an environment characterized by a high level of dynamism. Another main problem is related to the development of techniques for propagating social updates in an efficient way. Finally, even if data is no more stored on centralized servers, new privacy issues have to be solved, for instance detecting trusted nodes that may host the profile of off-line users.

This paper reviews the main solutions proposed in the literature to solve these challenges and presents a classification of the existing solutions. We describe in detail the DOSN research challenges, focusing, in particular, on the data management problem. First we introduce a classification of the main architectural solutions for DOSNs. Then, we introduce existing approaches for guaranteeing data availability, and the main techniques used to spread social content among the users of the social network.

Existing surveys or taxonomies [3,6,7] provide a detailed description of DOSNs, mainly from a privacy and security point of view. They address only partially other important issues, such as managing social data in a distributed environment, and do not provide a classification of existing proposals based on this characteristic. On the other hand, we focus on the data management problem in DOSNs and present current techniques for data availability, information diffusion, and data privacy in a distributed environment.

The rest of the paper is organized as follows. In Section 2, after a general overview of DOSNs, we propose a coarse classification of existing systems and we introduce the main research open issues. Section 3 presents the structures used to represent social data. Section 4 introduces current techniques to guarantee data availability and classify them, while Section 5 discusses and classifies the techniques used to spread social information. In Section 6 we describe data privacy in DOSNs. In Section 7, we present relevant examples of DOSNs. Section 8 discusses the limitations of current solutions and open problems and concludes the paper.

2. Decentralized Online Social Networks: overview and key research challenges

A general model for the decentralization of the services in an OSN is given in [8], where a DOSN is defined as a distributed system including a Social Network (SN), a Social Networking Service (SNS) and a Communication and Transport (CT) level. Using a network stack layering representation, SN is located on the top layer, SNS is the intermediate layer, and CT is the bottom one. The SN level provides the common social network functionalities, such as chat, mails, wall posts and/or tweets, etc..., while the SNS is usually implemented by a P2P network, i.e. a distributed network composed of a large number of distributed, heterogeneous, autonomous, and highly dynamic peers sharing their own resources (processing power, storage capacity files contents, etc...). The participants of the P2P network can act as a server and a client at the same time and their functionalities are accessible by other peers directly, without passing through intermediary entities. Finally, the CT level consists of Internet, mobile or opportunistic infrastructures that are used by the above levels to communicate.

DOSNs can be classified by considering the overlay connecting the peers of the users participating to the social network and, possibly, a set of external resources. As shown in Fig. 1, DOSNs can be classified in three categories. A first coarse grained distinction depends on whether external storage services are exploited. Solutions exploiting only the peers of the users participating to the social network can be further classified as:

- solutions exploiting a Distributed Hash Table (*DHT-based system*), whose nodes are those of the users participating to the social network, where the overlay connecting the peers is defined by the specific DHT topology. The DHT can be exploited both to store the social content and as an indexing service.
- proposals where peers are connected through a social overlay (SO) (*SO-based system*), where a logical connection between a pair of nodes corresponds to a friendship relationship. Solutions is this class may also exploit a DHT, generally only as an indexing service.

In those systems that exploit the DHT to store social content, data are stored encrypted in the DHT [9]. Since mapping of data to node storages is guided by an hash function, users can not control where data are stored. Other systems choose "trusted" nodes to store content replica and use the DHT only to index these replica.