Decentralizing Social Networking Services

Thomas Paul, Sonja Buchegger, Thorsten Strufe

Abstract Online Social Networks (OSN) of today represent centralized repositories of personally idenfiable information (PII) of their users. Considering their impressive growth they arguably are the most popular service on the Internet, both by technology savvy but even more by comparibly inexpert audiences, today. Being voluntarily maintained and automatically exploitable, they are a promising and challenging target for commercial exploitation and abuse by miscreants. Several approaches have been proposed to mitigate this threat by design. Removing the centralized storage, they distribute the service and data storage, to protect their users from a provider that has access to all the information users put into the system. This paper gives an overview of currently proposed approaches, and classifies them according to their core design decisions.

1 Introduction

Online Social Networks (OSN) are currently revolutionizing the way people interact, and are becoming de facto a predominant service on the web. The impact of this paradigm change on socio-economic and technical aspects of collaboration and interaction is comparable to that caused by the deployment of the World Wide Web in the 1990's.

Catering to a broad range of users of all ages and a vast difference in social, educational and national background, OSN allow even users with limited technical skills to publish Personally Identifiable Information (PII) and to communicate with an ease, sharing interests and activities with their friends or indeed anybody on the web. OSN contain digital representations of a subset of the relations that their users,

Thomas Paul and Thorsten Strufe
TU Darmstadt, Darmstadt, Germany, e-mail: `<lastname>@cs.tu-darmstadt.de`

Sonja Buchegger
KTH, Stockholm, Sweden, e-mail: `buc@csc.kth.se`

both registered persons and institutions, cultivate in the physical world. Centralized Social Network Services (SNS) manage, and offer online access to these OSN.

Adapted from the original definition in [3], an online social network can be defined as an online platform that (1) provides services for a user to build a public profile and to explicitly declare the connection between his or her profile with those of the other users; (2) enables a user to share information and content with the chosen users or public; and (3) supports the development and usage of social applications with which the user can interact and collaborate with both friends and strangers.

In centralized OSN all personal content is stored, at least logically, at a single location. This data store contains a very valuable collection of private information because detailed information about customers or potential new customers is very useful for the advertising industry. Centralized OSN need an operator to provide for the resources and to maintain this service, and hence this property is vital, since the primary way of financing current OSN is based on advertising business models. The idea to utilize this data treasure to gain money for the Social Network Provider is not far. The importance of this privacy exposure is underlined by the market capitalization of OSN providers, which ranges from 580 million US$ (acquisition of myspace through the news corp. in 2005) to 23 billion US$ (Facebook Inc, according to the investment of Elevation Partners in 2010)[1]. Even considering the commercial bodies that provide SNS to be trusted, hackers may be able to compromise their systems to gain access. Unsatisfied employees may abuse their access to the data, or even imprudent publication of seemingly anonymized data may lead to the disclosure of PII, as it has happened in the past[2]. In consequence, we consider the protection of PII in OSN as an emerging topic, which is currently not addressed by the providers in an appropriate way.

This article discusses approaches to decentralize the infrastructure of OSN's to avoid maintenace cost and the formation of data treasures. By taking away the single service provider, some privacy and performance problems of existing OSN can be solved as well.

2 Functional Overview of Online Social Networks

Social networking sites developed from early, simple online tools to manage personal and professional contacts to effective tools for sharing several kinds of information and content. Popular OSN, such as, e.g., Facebook, offer users even more services and applications, as third-parties are allowed to develop and plug their applications into the site. OSN hence have come closer to being full-fledged development and management platforms for social applications.

Even though each OSN is usually tailored to some specific use, the functional range of these platforms is essentially quite similar. Generally speaking, OSN func-

[1] `http://www.reuters.com/article/idUSTRE65S0CZ20100629`

[2] `http://www.nytimes.com/2006/08/09/technology/09aol.html`

tionality can be classified into three main types: The networking functions serve the actual purpose of OSN to foster social relationships amongst users within the virtual platform. In particular, they provide functionality for building and maintaining the social network graph. The data functions are responsible for the management of user-provided content and communications amongst the users. Their variety contributes to the enhancement of users' interaction and makes the platform more attractive.

2.1 Networking functions

OSN users can typically build their profiles and establish relationships with each other. The set of networking functions includes all functions that update the vertices and the edges of the social network graph. In particular, the OSN user invokes the profile creation function upon his or her registration on the OSN platform. This function adds a new vertex representing that user to the social network graph. Thereafter, with profile lookup the user can find other users, who are also represented via vertices. Through the call to the relationship link establishment function the user can set up a new relationship with some other user. This function typically sends notification to that user, who in turn can accept or ignore the request. If the user accepts the request then users are added to the contact lists of each other and a new edge representing their relationship is added to the social network graph. The OSN users can also encounter profiles for possible relationships by browsing though the contact list function, which is realized through the traversal along the edges of the graph. Additional networking functions can be used to remove vertices and edges from the graph, for example upon the deletion of a user's profile.

2.2 Data functions

OSN users can typically advertise themselves via their own profiles and communicate with each other using various applications like blogs, forums, polls, chats, e-mails, and online galleries. Here we point out the profile update function, which allows the OSN users to maintain details on their own profiles and provide fresh information to other users, who may call the profile retrieval function, and hence visit the profile. Communication amongst users via blogs and forums is typically implemented through a posting function, which inserts a block of information as an element into the main thread (sometimes called the "wall"). This block of information is not limited to plain text and can also contain videos, pictures, or hyperlinks. Updates to the profile or main thread, or a subset thereof, often are shown as a news feed on the main OSN page of connected users.

An OSN user willing to set up multimedia galleries typically calls the upload function, which transfers digital data from the user's device to the OSN database. In case of content depicting other users, the tag function can be used to create a link pointing to their respective profile. OSN users can typically evaluate content published by other users through the like or dislike functions. Using the comment function OSN users can articulate their point of view in a more explicit way. OSN users can also exchange personal messages. Here, in particular, basic, asynchronous offline communication (comparable to email) is implemented, and synchronous realtime communication between online users is offered in the form of chats. OSN users can send messages to individuals or to subgroups of users from their contact list. Additionally, users may create and join interest groups.

3 Decentralizing Online Social Networks

All users of centralized OSN request the service, and hence cause traffic at the social networking service provider. Growing up to serve millions of users, these central services naturally evolve to being bottlenecks. This problem has become increasingly apparent with the frequent down times and service break downs that users of highly popular OSN have experienced in the recent past[3]. Decentralized OSN avoid this disadvantage by distributing and making the stored data available from multiple locations. This inherently leads to a protection of the data from unintended centralized access and exploitation.

Decentralizing the service provision, however, raises a couple of ***requirements***. The distribution of OSN, which are catering for a broad range of users that frequently include large inexpert audiences, needs to be entirely transparent. Access to the data and functions needs to be provided through a single *integrating interface* which has to allow for easy publication, search, and retrieval of profiles and attributes. All *data related functions* of centralized OSN have to be provided in addition. The possibility to *reconstruct the social graph* of relations between the users finally has to be provided as well, in order to allow for simplified, often publish-subscribe-like communication, ease of access control, and publicly announcing real world friend- and other relationships. The distribution additionally must not lead to an interrupted *availability of data and services*, not even of parts thereof, even in face of the transition from dedicated servers to distributed resources. *Confidentiality* has to be met and *access* to each attribute *controlled*, even considering the lack of centralized management and control. Preserving the *privacy* of users and their data, even to an extent at which they are able to hide their participation inside the OSN completely, needs to be supported.

OSN in general may be decentralized at a different granularity. Integrating multiple commercial OSN [2, 8] and keeping chosen, partial data within the bounds of

[3] http://www.allfacebook.com/2009/08/facebook-downtime-issues/
http://www.pcworld.com/article/173550/facebook_outage_silences_150000_users.html
http://www.pingdom.com/reports/vb1395a6sww3/check_overview/?name=twitter.com\%2Fhome

different SNS represents a simple step towards decentralization. This approach removes the omniscient, commercial service provider with access to the overall set of PII of the users. It however introduces the role of the aggregator, which, even though only catering for a subset of all participants in the integrated OSN, again can collect complete knowledge about its users. To achieve better decentralization, the service provision can further be distributed, up to a granularity of a single service provider for each user. Proposed approaches at this level of decentralization can be divided into two different groups:

1. web-based decentralized OSN, and
2. Peer-to-Peer (P2P) OSN.

These groups represent differences on a rather high level, and hence are explained in more detail in the following sections.

3.1 Web-based decentralized OSN

Systems in the first group (mainly comprising of diaspora[4],[5] and “Friend-of-a-Friend” (FoaF) [11]) leverage on a distributed web server infrastructure. They require the acquisition of webspace or the deployment of additional web servers through their participants. Users then can publish their profiles much alike web pages in their own web space and locally manage access control rules to specifically allow retrieval of restricted attributes and resources to selected users. Web-like links to the profiles of other users are employed to represent the contacts list, and hence recreate the social graph.

The main *challenge* for these systems is their need for access to reliable web space, without which the profiles of the respective users are unavailable. Many, especially less tech savvy users experience major difficulties when being confronted with the task of setting up a web server themselves. Especially the task of reliably providing this service from home, including the configuration of home gateways, NAT, and firewalls, represents a serious obstacle. Renting web space, on the other hand either comes with the lack of being able to implement fine grained access control, and is quite costly in comparison to the existing free OSN, or does not help decreasing the complexity – and difficulties – of administrating them. This challenge of course generates a new business model, the provision of reliable, pre-taylored web space, including massive data aggregation at the provider and the resulting adverse consequences for the privacy of its users. A systematic challenge to these systems is the possibility to search for profiles of other users, like the proverbial long-lost-friend, since this is difficult to be implemented inside the systems. It rather has to be implemented in search engines, which again can gather knowledge about the users at a large extent.

[4] `http://www.joindiaspora.com`

[5] diaspora still is in the course of development and rather rapid changes. It is considered as it has been proposed during the time of this writing.

3.2 P2P OSN

The second group of systems [4, 5, 1, 7] harnesses the advantages of the well-known Peer-to-Peer principle in order to allow for the publication, search, and retrieval of profiles and their attributes, much alike the sharing, searching, and downloading in conventional P2P filesharing systems.

Challenges for P2P OSN are mainly caused by the different properties of file sharing vs social networking. P2P file sharing systems have been designed for the purpose of reliably distributing comparibly few, large, popular data objects (music files, movies). The automatic replication of these files during download led to an inherent load balancing, since more popular resources are downloaded, and hence replicated, more often. File sharing systems, however, only offer best-effort services, and the availability of less popular resources is all but guaranteed. Considering profiles in OSN exhibits a drastically different situation. Data in OSN consist of a profile for each user, each of which comprising a plethora of personal attributes. All these attributes of the large majority of users enjoy a very low popularity, but even though requested only very occasionally have to be kept available at all times. Owner replication, the provision of data at downloading parties, requires some registration of each resource (for the purpose of finding the replica), which becomes a difficult task considering the sheer numbers of single attributes. Another adverse property along the same lines is the fact that while the data in file sharing generally is accessible by anyone, access to private attributes of the profiles is restricted and they hence may not be replicated at arbitrary peers. Timing constraints are difficult to meet in P2P OSN: While users in file sharing are willing to wait even comparibly long intervals to download a complete movie or song, the user of an online social network expects the requested profile to be represented with very low delays. Even the user behavior poses a significant challenge for P2P OSN: The users of file sharing system are willing to stay online as long as it takes to download rather large files, whereas the users in P2P OSN will usually login, browse a few profiles, send a few messages, and log out after having been online for a couple of minutes, only; providing a reliabel peer-to-peer data service in this scenario causes serious obstacles for the system designers.

Peer-to-Peer OSN in conclusion also face serious challenges when striving at providing reliable social networking services.

4 Classifying Decentralized Social Networking Services

Analysing the two described groups of decentralized social networking services, they can be classified according to a few distinguishing characteristics. Their main target being the publication of profile information, while preserving the privacy of their users, they follow central design choices according to the four following properties:

1. The type of storage
2. The granularity of storage
3. The level of integration
4. Resource sharing incentives

The properties and classified groups are further explained in the following sections. Table 1 gives an overview of the analysed systems and their classification.

4.1 Type of Storage

Depending on the type of storage, the approaches can be classified into two groups. The first group, mainly consisting of the web-based approaches (diaspora and FoaF [11]) as well as Vis-a-Vis [10] leverage on dedicated servers. FoaF and diaspora on the one hand assume access to dedicated web space at which the profiles of users can be stored and retrieved. Vis-a-Vis on the other hand proposes to replicate to complete P2P software to a virtualized server in the cloud. Dedicated services, of course, come at an explicit, additional cost.

Likir [1], PeerSon [4] and Safebook [5] propose to only leverage on the local and shared resources of the P2P overlay. Leveraging on the rather unreliable storage services of other peers, who are subject to churn themselves, requires more sophisticated means of keeping the data available, which in turn causes a higher overhead and implicit cost, shared between the participants of the system.

LifeSocial [7] represents a hybrid approach. It implements a PAST [6] reliable P2P storage between the participating nodes, and additionally allows to acquire storage space at a dedicated server as premium services, for the purpose of guaranteeing the availability of data.

Approaches	**(1) Type of storage**	**(2) Storage granularity**	**(3) Level of Integration**	**(4) Ressource Sharing**
$diaspora$	web-based	complete	external services	premium services
$FoaF$	web-based	complete	external services	
$LifeSocial$	hybrid	split	stand alone	premium services
$Likir$	p2p	split	stand alone	
$Peerson$	p2p	split	external services	
$Safebook$	p2p	complete	stand alone	cooperation
$Vis-a-Vis$	dedicated	complete	external services	

Table 1 Classification and properties of the analysed systems.

4.2 Granularity of Storage

The granularity of remote storage ranges from replicating the complete service at the same place to storing each attribute at different places in the system.

The web-based approaches (diaspora, FoaF [11]), as well as Vis-a-Vis [10] and Safebook [5], bundle the complete service of delivering profiles. While the web-based approaches place the whole profile remotely in a single web space, Vis-a-Vis migrates the P2P software to a virtual server entirely. Safebook creates multiple replica of the complete profile, one at each of the profile owner's direct friends.

The remaining approaches (Likir [1], LifeSocial [7], and PeerSon [4]) split the profile into its attributes and replicate each attribute at different places. The complete profiles, some of which might be quite large in volume, are split, and comparibly small, single attributes are replicated in this case. The load of storing data for others consequently may be balanced more evenly, which, in turn, may lead to a reduced need to incentivize cooperation. However, it causes increased messaging overhead for the location and retrieval of each of the attributes, and eventually the complete profile.

4.3 Level of Integration

Implementations of SNS may either be self-contained, or integrating other services.

One group of approaches can be considered fully-fledged, stand-alone SNS, completely integrating the functionality and providing means to keeping data available anytime. These especially comprise of Likir [1], LifeSocial [7], and Safebook [5].

The other group of systems leverages external services for the replication and availability guarantees. Vis-a-vis [10] envisions to replicate the complete service to the cloud, which is expected to offer reliable availability, when the user is offline.

The first prototype implementation of PeerSoN [4] uses a third party DHT (openDHT [9]) as a lookup service to find content, and the SNS peers for storage. OpenDHT can be replaced by any DHT offering similar service (put, get, remove of entries) or by a self-contained peer implementation, combining the functionalities of storage and information administration in one system. PeerSoN, also envisions the option of using dedicated services that have high online probabilities, such as home routers or individual cloud storage, for users whose mobile or desktop resources are limited in extent and availability. Leveraging on external services, while potentially increasing the availability of data, comes at the cost of depending on them. Break downs and performance deficiencies have a direct impact on the operation of the SNS. Integration of commercial services, like cloud storage or computing most certainly cause additional cost.

These classes do not identify web-based decentralized SNS (diaspora, FoaF [11]) very well, since they are integrated into the web, rather than being stand-alone systems. They hence mainly comprise of a web page description scheme, and possibly an interface for their usage.

4.4 Resource Sharing Incentives

Implementing an integrated SNS with replication that cannot rely on external storage systems results in the need to incentivise service providers to actually store the replica, to keep them available, and to eventually deliver them to requesting users.

Different incentive schemes have been proposed in literature that could potentially be integrated and simply utilized. However, none of the existing approaches follows this strategy. While not all of the proposed approaches actually consider this need, the chosen solutions can be generalized to two different types: financial and social incentives.

Some solutions, like, e.g., diaspora, or LifeSocial [7], consider the possibility of offering payed premium services through the system provider, which hence enjoys *financal incentives*. These premium services would comprise a centralized replication of the premium profiles for a fee, in order to keep them available at all times.

Safebook [5] takes another approach of considering *social incentives*: since friends of a user generally are trusted and believed to cooperate, the main profile information of each user is replicated to all their friends' devices. Complex, additional networked structures, the "matryoshkas", are created for the purpose of hiding the friend relationship from other participants, and they are optimized to increase the chance of locating the profile replicas. However, the availability of profiles may not be guaranteed, if the number of friends of a user is too low, or in case that all of a user's friends concurrently are offline.

5 Conclusion

The information we reveal about ourselves online has changed both quantitatively (more volume) and qualitatively (increasingly personal), especially over the last decade. Both trends are best exemplified by online social networks. Web services based on an advertising business model, in parallel, have gained market share. Given that, the model of an attention economy that such business models (based on advertising) are building on, is by necessity one of scarcity. Our attention as humans is limited by the hours in a day. Given this rather hard limit in how many effective advertisement-based services can co-exist, advertisement has to be targeted to user interests in order to get enough click-through and ”eye-balls” to support the service. The more targeted the advertising, the more personal and demographical information about prospective customers is needed. This model thus renders information about users more valuable, resulting in an incentive for service providers to gather even more personal information.

We therefore increasingly observe that service providers, especially those of OSN, are pushing the boundaries of extracting personal information from users. Online social network providers have been attracting users to share an increasing range of personal data that is shared along the lines of friends, friends of friends, other users, and, finally, anybody on the Internet. Despite outcries about privacy violations and difficulties of configuring privacy setting preferences, this trend continues. Following pressure from users, Facebook, for example, recently changed the way privacy settings are made in an effort to make it easier for users. Their default settings, though, are quite far from what one would expect as privacy preserving. The typical user who relies on default settings thus gets privacy settings that share vital personal information with everyone on the Internet.

In an effort to preserve user privacy while keeping useful features offered by online services, such as social networks, there is increasing research activity proposing to go from centralized provider-based models toward a community-driven decentralized approach, based on peer-to-peer networks. In this paper, we discussed several of these approaches and classified them according to design decisions such as whether they are web- or p2p based, whether they integrate third-party services, how storage is provided, and others. This is a current snapshot of research projects for decentralized SNS and we anticipate more and different approaches in the near future. Our survey and classification serves as a first step toward distilling best practices from different approaches to decentralizing SNS. Over time, given lessons learnt from implementations, experiments, and hopefully even user adoption, such classifications and evaluations enable designers of decentralized SNS to leverage results from others and build privacy-preserving SNS that exhibit desireable features suach as low overhead, high availability, and reliablity.

References

1. Aiello, L.M., Ruffo, G.: Secure and Flexible Framework for Decentralized Social Network Services. In: SESOC 2010: IEEE International Workshop on SECurity and SOCial Networking (2010)
2. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: ACM Internet Measurement Conference (2009)
3. Boyd, D., Ellison, N.B.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication (2007)
4. Buchegger, S., Schiöberg, D., Vu, L.H., Datta, A.: PeerSoN: P2P social networking - early experiences and insights. In: Workshop on Social Network Systems (2009)
5. Cutillo, L.A., Molva, R., Strufe, T.: Safebook: a Privacy Preserving Online Social Network Leveraging on Real-Life Trust. IEEE Communications Magazine **47**(12), 94 – 101 (2009)
6. Druschel, P., Rowstron, A.: Past: A large-scale, persistent peer-to-peer storage utility (2001)
7. Graffi, K., Podrajanski, S., Mukherjee, P., Kovacevic, A., Steinmetz, R.: A distributed platform for multimedia communities. In: International Symposium on Multimedia (2008)
8. Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., Farrell, S.: Harvesting with sonar: the value of aggregating social network information. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1017–1026 (2008)
9. Rhea, S., Godfrey, B., Karp, B., Kubiatowicz, J., Ratnasamy, S., Shenker, S., Stoica, I., Yu, H.: OpenDHT: a public DHT service and its uses. In: SIGCOMM (2005)
10. Shakimov, A., Cox, L., Varshavsky, A., Caceres, R.: Privacy, cost, and availability trade-offs in decentralized osns. In: Workshop of Online Social Networks (2009)
11. Yeung, C.M.A., Liccardi, I., Lu, K., Seneviratne, O., Berners-Lee, T.: Decentralization: The future of online social networking. In: W3C Workshop on the Future of Social Networking Position Papers (2009)