

An Analysis Grid for Privacy-related Properties of Social Network Systems

Regina Paiva Melo Marin, Guillaume Piolle, Christophe Bidan

► **To cite this version:**

Regina Paiva Melo Marin, Guillaume Piolle, Christophe Bidan. An Analysis Grid for Privacy-related Properties of Social Network Systems. SOCIALCOM 2013, Sep 2013, Washington D.C., United States. pp.520-525, 10.1109/PASSAT/SocialCom.2013.79 . hal-00908339

HAL Id: hal-00908339

<https://hal-supelec.archives-ouvertes.fr/hal-00908339>

Submitted on 22 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Analysis Grid for Privacy-related Properties of Social Network Systems

Regina Marin, Guillaume Piolle and Christophe Bidan

SUPELEC

Cesson-Sévigné, FRANCE

Email: {regina.marin, guillaume.piolle, christophe.bidan}@supelec.fr

Abstract—Social Network Systems (SNSs) are the predominant kind of web service around the world. They attract many users seeking popularity, entertainment and network building, along with ease of use. Most current SNSs are based on centralized designs, which are less likely to improve privacy since there is a single and central authority with exclusive administration control over user information. Many proposals have been introduced that work towards decentralizing the infrastructure support in order to enhance privacy in SNSs. However, designing decentralized social network systems (DSNS) driven by privacy is a hard task because privacy is impacted by most design choices. This paper proposes a multi-criteria analysis grid designed to evaluate several properties of SNSs related to privacy trade-offs. Based on the analysis grid result, this paper also presents the application of lattice-based tools to classify and visualize social network systems in privacy-related hierarchies.

Keywords—privacy; social network systems; evaluation; lattices;

I. INTRODUCTION

Privacy in our online society has proven difficult to achieve. Many concerns about data privacy are related to the information sharing occurring in Social Network Systems (SNSs). A widely accepted definition of SNSs was provided by Boyd and Ellison [1], who describe social network sites as “*web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.*”

According to Nielsen [5], SNSs are the most visited worldwide websites and gather a huge amount of sensitive information in data centers. As a consequence, a variety of privacy shortcomings have arisen in current SNSs, mainly because of their centralized infrastructure. As an example, SNSs service providers usually have unlimited access to users’ information and total control over its retention and use. Therefore, users are not left with much control over how their personal data is collected, used and disseminated. Aïmeur, Gambs and Ho [9] define privacy-enhanced SNSs as SNSs with: (1) *privacy awareness and customization*, to help users take informed decisions about information sharing, (2) *data minimization*, to ensure that only necessary information is collected and processed, and (3) *data sovereignty*, to ensure that users remain in control of their data. These

very general properties still need to relate to the design choices leading to them. Among them, decentralization seems to us a promising and efficient way to put users in control and ensure the privacy of their information. Indeed, decentralization tends to be seen as a transfer of control and services from service providers to users, giving privacy protection a leading role in the specifications.

Many alternative SNSs use decentralized data and/or services in order to enhance privacy by keeping personal data on users’ devices and maintaining policy enforcement out of the control of a single entity. However, the absence of a central server introduces certain difficulties in managing the infrastructure support. In particular, one should provide a balance between privacy protection on the one hand, and security, data availability and usability on the other hand. In fact, each solution focuses on decentralizing some specific design points according to a prioritization of trade-offs based on the designer’s preferences. Therefore, the emerging challenge is to evaluate, compare and classify the existing SNSs according to their core design choices, which are not obvious nor direct for the SNS designer.

If we accept that decentralization of data and services enhances the general level of privacy, then it will be useful to be able to evaluate a given solution according to its actual design choices, in terms of decentralization. We believe that such a measure would be a relevant rating of its privacy-friendliness. In this perspective, we propose to classify privacy-related properties in a multi-criteria analysis grid based on degrees of decentralization. We think that such a grid can be useful in a *Privacy by Design* methodology to develop SNSs. Indeed, based on the grid, it is possible to organize privacy-related design choices in a lattice and benefit from the associated formal structure to compare and classify SNSs based on sets of technical options.

The paper is organized as follows. Section II presents the context related to social network systems. Section III describes the privacy-related properties we find relevant to SNS design. Section IV presents the multi-criteria grid analysis based on degrees of decentralization and its application to social network systems. In section V we present the lattice structure used to systematically organize sets of privacy-related design choices into hierarchical structures. Section VI compares our approach with related works. We finally offer a conclusion and propose future research tracks.

II. SOCIAL NETWORK SYSTEMS

A. Centralized Social Network Systems

SNSs are a well-established global market phenomenon. They provide various benefits to users, such as availability of web space to create their profiles and the ability to establish bonds with their friends, relatives and acquaintances. All these features are specifically designed to be easily usable, even by beginners. In addition, a SNS is a central place to benefit from several applications offered by the service provider and third parties, such as music and games. Common examples are Facebook¹, Twitter², LinkedIn³ or MySpace⁴.

Such centralized systems are relatively easy to design and allow to set up services useful to both parties. However, centralization of user data and usage control is a significant threat to users' privacy, as made obvious by several recent abuses⁵. From the point of view of privacy policy management, the central authority imposes a global policy, but every single user may also have his own one. However, these policies are not on the same level and the central authority may constrain user policies, which usually do not even apply to it. Moreover, recent studies have also shown that few users actually read privacy policies, and that policies published by SNS providers typically require college-level reading skills [2]. Beyond that, these policies may not reflect adequately their actual use of user data [12]. User privacy policies are often implemented in SNSs as access control policies. Yet, there are usually no access control policies between users and the service provider. As a consequence of the (often) unlimited access to user information, SNSs providers may perform data mining and profile analysis for targeted advertisement, and usually sublicense this valuable information in order to generate income.

B. Decentralized Social Network Systems

To overcome the issues raised by centralized control, several alternative platforms have been proposed. These systems introduce distinct degrees of decentralization to enhance privacy. The Diaspora network⁶ is an open-source SNS project based on a decentralized architecture using servers called "PoDs". Each public PoD can be connected with a set of clients called "seeds". PoDs are in charge of profile data storage (in their local database), contact search (by querying other PoDs) and information retrieval.

PeerSoN [14], [8] is a decentralized SNS using an external P2P infrastructure called OpenDHT. An OpenDHT overlay can be seen as a set of super-peers, which will provide

lookup services and maintain profile information in a distributed hash table, in the form of key-value pairs. Each peer can use the services provided by the super-peers to locate the peer in charge of a given piece of data and directly communicate with this peer to retrieve the data.

PrivacyWatch [9], [7] is a hybrid (partially decentralized) approach that focuses on the trade-off between privacy and usability. In this context the client privacy manager (CPM) is a browser plug-in which helps users to set their privacy level and privacy preferences. After that, the mail server is used by the CPM to create an e-mail account used for key sharing. The SNS provider is used for searching friends and storing encrypted personal information.

Safebook [4], in a nutshell, is a decentralized SNS, using a structure called *matryoshka*, which can be described as a set of concentric rings of nodes built around each user node in order to provide trusted data storage, profile data retrieval, communication obfuscation and anonymization through indirection. A lookup service is provided in the P2P architecture for finding entry-points for matryoshkas, using pseudonyms provided by the trusted identification system.

FOAF (Friend-Of-A-Friend) [10] is an attempt to improve data interoperability in SNSs. It uses an ontology to describe people by their attributes (`foaf:person` elements) and their social connections (`foaf:knows`). Users are put in control by allowing them to decide which server will store their FOAF profile.

Finally, SuperNova [13] focuses on data availability in decentralized SNSs. Super-peers provide data storage and lookup services to peers that do not have enough friends in the SNS. Storekeepers do the same but only for their friends. Thus, to ensure data availability, each peer can ask its friends to become storekeepers for its data. When this peer is down, its data can still be accessed through its storekeepers.

III. PRIVACY-RELATED PROPERTIES FOR SNSs

The cited works aim to improve different aspects of user privacy, by focusing on different design choice points. Here we have identified the main core design choices to be integrated into the design of an SNS through the following properties. We conceived these properties to be a synthesis of the trade-offs in the existing approaches. In our opinion, bringing more decentralization to the different aspects of SNS design is likely to improve the general level of privacy, by avoiding the too significant influence of a central authority. Therefore, we choose to evaluate privacy-related properties with respect to the degree of decentralization assigned to it in the SNS design.

A. Degree of decentralization

We consider three degrees of decentralization: (i) *centralized (C)*, (ii) *decentralized (D)* and (iii) *fully decentralized (FD)*, which correspond to the columns in the grid in section IV.

¹<https://www.facebook.com/>

²<https://twitter.com/>

³<http://www.linkedin.com/>

⁴<http://www.myspace.com/>

⁵<http://www.europe-v-facebook.org/>

⁶<http://diasporaproject.org/>

Centralized SNSs have a strongly hierarchical structure. There is a single and central authority with exclusive administration control. Centralized SNSs have a star network topology, meaning that all peers are directly connected with the central authority. This is typically implemented as a client-server organization, the central authority being in charge of communication routing, friend search and content retrieval on behalf of the peers.

A first step in decentralization is to avoid the unicity of this central authority and to allow for local, autonomous authorities to emerge. Such an organization correspond to the **decentralized** category in our classification. These systems have a hybrid network topology including a set of autonomous authorities with local administration control (sometimes known as “super-peers”), as well as ordinary peers.

The next step is to build a **fully decentralized** SNS, where each peer can be seen as a punctual authority. Neither peers nor the network itself are organized in a hierarchical structure. All peers are equals in terms of service providing and control over data. Interactions are usually implemented through direct communication between peers.

B. Description of Privacy-related Properties

Privacy-related properties correspond to the rows in the grid in section IV. To each of these properties corresponds a gradation along the decentralization scale previously introduced.

1) *Architectural Services*: They cover the main services provided by the SNS, such as search, data retrieval and communication. **Search** is the mechanism to locate data and peers in SNSs. **Data retrieval** is the mechanism through which data is exchanged among entities (peers, service provider and third parties). **Communication** is how data is transmitted among entities.

- Search
 - *C*: Only the central authority is in charge of searching friends/content for all peers.
 - *D*: A given set of autonomous authorities are in charge of searching friends/content for all peers.
 - *FD*: The set of all peers are in charge of searching friends/content for all peers.

2) *Storage*: It describes how information is kept in the system. One important feature related to storage is data availability. Often, SNSs apply replication techniques to make data redundant. We propose three properties: **Storage space** tells us where peer data is stored, **replication** indicates which entity is in charge of replicating profiles and resources, and **data suppression** specifies which entity has the power to delete data from the system (for instance, when a user closes their account).

3) *Security Aspects of Privacy*: They correspond to the mechanisms used to protect data confidentiality and integrity

as well as peers’ identities and activities. Most privacy regulations require that personal information be kept secure. To characterize these properties, we first introduce an attacker model which will allow us to define our three degrees of decentralization in the context of security and privacy. In our attacker model, each attacker is able to fully compromise one or several entities in the system, and its aim is to affect all the peers of the SNS with respect to a given property. For a **centralized** property, the attacker need only to compromise the central authority in order to affect all peers. For a **decentralized** property, he must compromise a given set of autonomous authorities in order to affect all peers. Finally, for a **fully decentralized** property, the attacker must compromise all peers in order to affect all peers.

Two properties relate directly to encryption. They tell us which entity controls encryption and decryption of data, in the case of **data encryption**, or of communications, in the case of **traffic communication encryption**.

- Data encryption / Traffic communication encryption
 - *C*: Only the central authority must be compromised in order to decrypt data of all peers.
 - *D*: A given set of autonomous authorities must be compromised in order to decrypt data of all peers.
 - *FD*: The set of all peers must be compromised in order to decrypt data of all peers.

The following four properties are more specifically about privacy protection. **Anonymity** measures the capacity of a peer to perform an action within the SNS without disclosing its identity. One should note that this classification assumes the existence of a trusted authority either as a central authority for *C* or as a set of autonomous authorities for *D*. Therefore, *C* and *D* have a weak notion of anonymity and only *FD* is able to provide anonymity *stricto sensu*. **Pseudonymity** measures his capacity to perform an action within the SNS without disclosing its identity, and still be accountable for that action. **Unlinkability** measures the impossibility to establish correspondence between two independent and different actions performed by the same peer. In the SNS context, **unobservability** means the capacity of a peer to perform an action without others being aware of these actions.

4) *Privacy Policy Management*: It encompasses policy administration and policy enforcement. The **policy administration** property describes which entity is in charge of the definition and modification of policies, whereas the **policy enforcement** property tells us at which level the privacy policy is enforced. Both properties relate to two kinds of policies: **system policies** and **peer policies**. The system policy applies to the whole platform and governs the rights of the SNS provider, when it exists. Peer policies regulate privacy preferences among peers. The latter can be more or less rich and expressive depending on the systems:

peer policies can range from imposed, system-wide rules to individually negotiated agreements between pairs of peers.

IV. THE MULTI-CRITERIA ANALYSIS GRID

We now propose to organize the privacy-related properties previously described in a two-dimensional grid: each line corresponds to one property (belonging to one of the four groups detailed in section III-B), each column to a degree of decentralization. The degree of decentralization has been chosen as the evaluation criterion since we believe that the distribution of services and data has a significant impact on the global level of privacy of the system. In this respect we adopt the following scale: *Unknown* (?) means that there is no available information in the SNS specification; *Nonexistent* (0) means that the privacy-related property is explicitly not addressed or not implemented in the SNS; *Centralized* (C), *decentralized* (D) and *fully decentralized* (FD) are as defined in section III-A.

We apply our approach to seven SNSs and present the result in table I. For the sake of brevity we only detail the analysis of one examples, namely *Facebook*, the largest and most successful centralized social network in the world, having more than one billion users.

Architectural Services provided by the SNS to users are mainly based on a centralized architecture where search, communication and information retrieval services are operated by a central entity, the service provider at Facebook, and only the result is provided to users on the client side.

Storage is centralized in Facebook’s cluster of around 180,000 web and database servers. Facebook replicate the complete user profiles across their data center. Data suppression does not seems to be implemented, because Facebook apparently remains with users’ data for an indetermined time, arguing safeguard against legal measures.

Security Aspects of Privacy rely on traffic communication encryption using SSL/TLS in order to provide security of the communication between the user’s browser and Facebook’s servers. Regarding this property, if Facebook servers are compromised, then the attacker will be able to decrypt all further traffic⁷. Data encryption is marked as nonexistent, because Facebook itself does not provide users with options to encrypt their data. Anonymity property is marked as nonexistent, since any communication is linked to a user’s personal account, itself based on a real-life identity (per the Facebook terms of service). Pseudonymity is also marked as nonexistent, for the same reason. Regarding unlinkability and unobservability, it must be noted that some actions (like people search or profile consultation) are visible and linkable only by Facebook as the central authority, while other (like status updating or public messaging) are visible and linkable by other users. Given the limited level of granularity we

⁷Of course, if one trusted Certification Authority (CA) is compromised, then it allows for man-in-the-middle attacks. This is true for all platforms relying on SSL/TLS and is left outside the scope of this analysis.

	Facebook	SuperNova	Diaspora	PrivacyWatch	PeerSoN	Safebook	FOAF
	? 0 C D FD	? 0 C D FD	? 0 C D FD	? 0 C D FD	? 0 C D FD	? 0 C D FD	? 0 C D FD
Privacy-related Properties							
Architectural Services							
Retrieval	x	x	x	x	x	x	x
Communication	x	x	x	x	x	x	x
Search	x	x	x	x	x	x	x
Storage							
Storage Space	x	x	x	x	x	x	x
Replication	x	x	x	x	x	x	x
Data Suppression	x	x	x	x	x	x	x
Security Aspects of Privacy							
Data encryption	x	x	x	x	x	x	x
Traffic encryption	x	x	x	x	x	x	x
Anonymity	x	x	x	x	x	x	x
Pseudonymity	x	x	x	x	x	x	x
Unlinkability	x	x	x	x	x	x	x
Unobservability	x	x	x	x	x	x	x
Privacy Policy Management							
P.A. System policy	x	x	x	x	x	x	x
P.A. Peer policy	x	x	x	x	x	x	x
P.E. System policy	x	x	x	x	x	x	x
P.E. Peer policy	x	x	x	x	x	x	x

Table I
THE MULTI-CRITERIA ANALYSIS GRID APPLIED AT SNSs PROPOSALS.

have chosen in our analysis, we must then conclude that unlinkability and unobservability are of the centralized kind in Facebook.

Privacy Policy Management in Facebook is centralized for anything (administration and enforcement) regarding the system access control policy, imposed by the contractual terms of Facebook. Users also have the ability to set up what we have called a peer policy, also focused on access control. More specifically, users can categorize all of their contacts in groups sharing the same access rights. Furthermore, users can specify which posts and photos the audience may access based on the following presets: “public”, “friends”, “custom”, “close friends”, “family”, “acquaintances” and “only me”. However, the policy of each user is stored

storage [15]. Thus, they identify SNSs such as FOAF and Diaspora, that use trusted servers to provide these features, and those that are based on P2P systems, such as Safebook and PeerSoN. However, they do not take into account the security aspects of privacy nor privacy policy management.

Thus, the aforementioned works do not cover the complete set of properties present in our approach for evaluating the level of privacy of SNSs. Our taxonomy is based on the degree of decentralization of all the privacy-related properties, in relation to architectural services, storage, security and policy management. Furthermore, using lattices makes it possible to compare and identify which SNS “scores best” with regards to each specific property.

VII. CONCLUSION AND FUTURE WORKS

In this paper we have examined the concerns arising from centralized social network systems, as well as the current prominent approaches that try to overcome these problems through decentralization of SNS data and services. The efforts engaged to protect user data in decentralized SNSs aim at keeping the data with the users, on their personal devices, and at developing SNSs using *privacy by design* principles.

We have chosen to put a stress on various design properties which can be set at various degrees of decentralization, thus impacting the overall privacy level of the application. We have organized these properties, all related to privacy issues, in several families: architectural services, storage, privacy policy management and security aspects of privacy. Based on the hypothesis that avoiding central authorities limits the risks of abuse, we have developed a multi-criteria analysis grid to analyze and evaluate SNSs in this respect. Algebraic lattice theory can then be applied to this grid, allowing to build a comprehensive structure aimed at identifying which SNS performs better with respects to a given property, and more generally at comparing SNS platforms in terms of privacy protection. Using the proposed lattice structure, it is possible to classify, evaluate and visualize different SNSs within a partial hierarchy based on lattice chains and levels. We believe such a graphical and computational tool to be a useful and usable contribution to *privacy by design* techniques, allowing SNS designers in the specification phase to distinguish current best practices and to find out how to improve them for the sake of privacy.

Future works include enrichment of the grid with properties specifically linked to the privacy policies themselves, especially in terms of expressivity. We believe that this is another dimension along which it would be interesting to compare SNSs. Another possible track is the development of software components dedicated to the achievement of a given level of decentralization for a set of given properties. Such modular and reusable software, deeply linked to our analysis tools, could also find a place in a *privacy by design* conception framework.

Acknowledgements: This work has been partially funded by the grant “ARED Presodis” of the Bretagne region.

REFERENCES

- [1] D. Boyd and N.B. Ellison, Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*. 1–2, 2007.
- [2] L.F. Cranor, P3P: Making Privacy Policies More Useful. *IEEE Security and Privacy*. 50–55, 2003.
- [3] E. Bertino and C. Brodie and S.B. Calo and L.F. Cranor and C. Karat and J. Karata and N. Li and D. Lin and J. Lobo and Q. Ni and P.R. Rao and X. Wang, Analysis of Privacy and Security Policies. *IBM Journal of Research and Development*. 3:1–3:18, 2009.
- [4] L.A. Cutillo and R. Molva and T. Strufe, Safebook: A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust. *IEEE Communication Magazine*. 94–101, 2009.
- [5] Nielsen Holdings, State of the Media: The Social Media Report, <http://blog.nielsen.com/nielsenwire/social/>.
- [6] B.A. Davey and H.A. Priestly, Introduction to Lattices and Order. Cambridge University Press, UK, 1990.
- [7] A. T. Ho, Towards a Privacy-Enhanced Social Networking Site. PhD thesis, Montreal University, 2012.
- [8] S. Doris, A Peer-to-peer Infrastructure for Social Networks. PhD thesis, TU Berlin, 2008.
- [9] E. Aïmeur, S. Gamba, A. Ho, Towards a Privacy-Enhanced Social Networking Site. In: International Conference on Availability, Reliability and Security, pp. 172–179. IEEE Computer Society, Los Alamitos, 2010.
- [10] C.M. AuYeung and I. Liccardi and K. Lu and O. Seneviratne and T. Berners-Lee, Decentralization: The Future of Online Social Networking. In: W3C Workshop on the Future of Social Networking, Barcelona, Spain, 15–16, 2009.
- [11] Q. Ni and E. Bertino and J. Lobo, An Obligation Model Bridging Access Control Policies and Privacy Policies. In: 13th ACM Symposium on Access Control Models and Technologies, New York, USA, 133–142, 2008.
- [12] P. Anthonysamy and A. Rashid and P. Greenwood, Do the Privacy Policies Reflect the Privacy Controls on Social Networks?. In: The Third International Conference on Social Computing (SOCIALCOM), 1155–1158, 2011.
- [13] R. Sharma and A. Datta, SuperNova: Super-Peers Based Architecture for Decentralized Online Social Networks. In: Fourth International Conference on Communication Systems and Networks (COMSNETS), 1–10, 2012.
- [14] S. Buchegger and D. Schiöberg and L.H. Vu and A. Datta, PeerSoN: P2P Social Networking - Early Experiences and Insights. In: Second ACM Workshop on Social Network Systems, 2009.
- [15] T. Paul and S. Buchegger and T. Strufe, Decentralizing Social Networking Services. In: International Tyrrhenian Workshop on Digital Communications, 1–10, 2010.