Defamation-Free Networks through User-Centered Data Control

Nadim Sarrouh, Florian Eilers, Uwe Nestmann, and Ina Schieferdecker

Technische Universität Berlin, Germany
{n.sarrouh,f.eilers,uwe.nestmann}@tu-berlin.de,
ina.schieferdecker@fokus.fraunhofer.de

Abstract. Existing online social networks hardly care about users' privacy rights. In particular, they do not permit users to keep control over "their" data. By "their" data, we denote data that refers to the respective user as an identifiable object within (textual, audio, image or video) media. The well-known concept of "usage control" employs a usage rights' perspective (e.g. DRM), but it does not explicitly deal with privacy. In this paper, we instead propose the concept of "data control", which exactly focusses on privacy rights and therefore employs a control rights' perspective. Based on data control, we propose a defamation-free network (DFN) in which control rights are not only manifest and visible, but can also be exercised. We examine the main usage scenarios of such a network, and discuss the possible approaches for implementing it. Finally, we sketch a solution with an underlying P2P architecture and highlight the basic technological challenges and requirements.

Keywords: Privacy, Social Networks, Usage Control, P2P, Web of Trust.

1 Introduction

Today it is common for employers to search the web for crediting or discrediting information about prospective employees. *Google*'s new mobile phone "Nexus One" proposes applications, through which users may get all the online information available concerning the person that you took a picture of, using the internal camera. Google-CEO Eric Schmidt states, not as ironically as one would hope: "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."[1]

This kind of user transparency is undesirable in many cases. It is obvious that the liability for these circumstances cannot simply be seen with the data storage or search providers such as *Google*. Users themselves have to become sensible concerning their own privacy. They have to be made aware of possible dangers of unconditional personal exposure on the web. However, apart from these social aspects, the technology itself should provide better means not only to raise awareness but also to provide tools to deal with threatening exposure or defamation.

[1] For this statement watch Eric Schmidt on privacy on *Youtube*:
`http://www.youtube.com/watch?v=A6e7wfDHzew\&feature=player_embedded`
Last checked: 25.03.10.

J. Cuellar et al. (Eds.): STM 2010, LNCS 6710, pp. 179–193, 2011.

Several existing online services such as “DeinGuterRuf.de”[2] offer to help the user with finding and erasing this kind of data from the Internet on a commercial basis. These companies enforce the erasing of the data and relieve the user of the stressful and complicated communication with responsible website owners. If the Web site owner does not yield to the demand, legal steps may be taken in most countries in order to force him to delete this data. However, considering the distributed content-centric architecture of the web, there is up to date no technical solution to enforce those rights at an infrastructural level.

We believe that, considering these rising challenges of privacy, it is necessary to develop a new infrastructure in which the user not only has control over the distribution and usage of user-owned data (usage rights) but also over associated data (data that identifies the user in some way) posted by others (privacy rights). To this end, it is essential to investigate new technologies that enforce rights of the individual at a technical level.

Within this paper, following the summary of related work in section 2, we define data control in this privacy context in section 3. We derive four levels of data control and examine the main scenarios that an architecture with privacy control would have to stand up to. We conclude the third section with a list of basic technological requirements for implementing such an architecture. In section 4 we discuss the suitability of centralized and decentralized approaches. Finally, in section 5, we propose the development of a defamation-free P2P online social network and sketch the main basic challenges of such a Defamation-Free Network (DFN) before we come to our conclusion and outline future work in section 6 and 7.

2 Related Work

The nearest approach to our proposal is Castelluccia’s “Owner-Centric Network” [4]. He considers a novel network architecture in which users may track where they stored data themselves, so that it may easily be retrieved and if necessary modified or deleted by the owner. Users would then be able to control the flow of their own data and access control mechanisms would prevent other users to download or change this data. However, Castellucia’s approach does not consider any degree of control over personal data, which is spread by others and therefore does not satisfy our requirements for enhanced data control.

Research concerning P2P social networks has been lately very popular. The two approaches most suitable to our data control requirements are *Safebook* [5] and the *Peerson Project* [2]. Among these two platforms, the *Safebook* platform seems slighty further developed; therefore, it also looks more promising as a potential basis for an implementation of a DFN. For a concrete solution proposal see section 5.

Usage control concepts, platforms and enforcement mechanisms exist in a variety too big to cover in this paper. The most common model for usage control is the *UCONabc*-model [9], which proposes a novel view on data objects,

[2] `http://deinguterruf.de/` Last checked: 25.03.10.

consisting not only of the data itself, but of usage rights, obligations and conditions. Several researchers have tried to implement this concept or a variation of it. Pretschner et al. invented a policy language and formal means to express usage control mechanisms in a verifiable way. Based on these formalizations they have proposed a usage control architecture based on an *X11* and data flow tracking [11]. Alam et al. suggested a *SElinux* with *mandatory access control* (*MAC*) in order to enforce policies concerning data [1].

3 Data Control

In this section, we provide our view of data control. We start by pointing out the difference to the concept of traditional usage control and formulate an approach for possible data control levels. Finally, we motivate which data control level we want to enforce with our conceptual architecture.

3.1 Data Control vs. Usage Control

The concept of usage control consists of data-object specific authorization and access monitoring and enforcement mechanisms, which are to enable the constraining of future usage of data after it is released from its own environment into a different control domain. In order to get data from the data provider negotiations concerning the usage of this data take place. These negotiations result in a usage policy which is transferred together with the data, assuming the client is indeed able to enforce this policy through a local usage control platform. The usage control platform transfers the policies and (by monitoring the access and usage of the data) ensures that its constraints are enforced properly[10].

This concept is relevant for commercial online data distribution, e.g. in *digital rights management (DRM)*. An example scenario could be: “viewing this movie is permitted 5 times only and will be denied after 48 hours”. Also medical applications especially considering Electronic Health Records may benefit from usage control mechanisms. Hafner et al. propose a usage control architecture in order to provide patients with the means to enforce their privacy in scenarios like: “Access to a medical record is allowed for 5 times only and should last for 48 hours, after its first access”[8].

As opposed to usage control’s provider-consumer scheme, we propose the concept of data control. Just like usage control this notion is a collection of mechanisms. However, it does not exclusively deal with usage rights, but focusses on privacy rights. It provides mechanisms and tools that allow control over personal data spread throughout a network. Users have the possibility to get an overview of data that identifies them no matter if published by them or by others. They may even get the rights to unpublish it. We point out that such rights do not only apply to user-owned data. Associated data published by others which involves the user or in the worst case defames or discredits him, has to be controlled in the same manner.

The data control concept is relevant for personal use, but also suitable for groups or cooperations since actors do not have to be natural but can also be

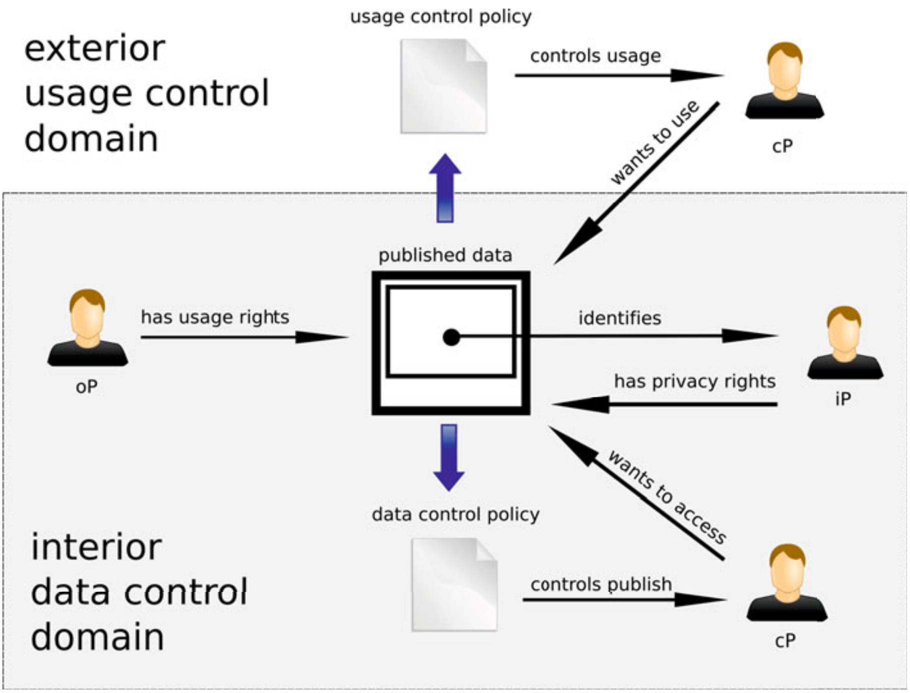

Fig. 1. Data control and usage control perspectives

legal bodies. One basic scenario would be: "How can I prevent that my employer gets hold of compromising data about me?"

Summarizing, the main differences of usage control and data control are based on different perspectives. While usage control is content-centered and focuses on the enforcement of usage rights, after data has been delivered from the owners domain to another environment, user-centered data control focusses on a user's privacy rights inside a social network and delivers tools and means to enforce his privacy inside this domain. Usage control and data control are therefore complementary concepts. Figure 1 visualizes a scenario where usage control is used as an extension of data control. Users are categorized into three roles: Owning persons (oP) may enforce usage and privacy rights on the data they own. Persons identified in published data (iP) may enforce their own privacy rights. Consuming persons (cP) may want to view or use the data and to do so must accept the terms of data control or usage control policies.

We identify four levels of data control, focusing on different aspects of data:

Data Control Level I. This basic level deals with the bare linking of data to users. Links are publicly visible associations of data to users, normally made manually by the user himself or by his friends. This level describes the possibility to get an overview over which data in the network is linked to the user's profile and the ability to unlink it. (e.g. Facebook photo links). The main motivation of establishing this data control level would be the argument that only data that is linked to a user can be found by using a name, dates, etc. Therefore it may be enough to delete these links in order to establish control over defaming or discrediting data.

Data Control Level II. Like level I, but including the possibility to not only overview data that is linked to a user's profile but also to find unlinked data associated to them. Associations fundamentally differ from links: They are automatically generated relations from data to users and are not publicly displayed. The only person aware of associations to data is the actual associated user. To get an overview of these associations, the infrastructure has to be equipped with content recognition algorithms. The content recognition would be triggered during the publishing of the data. This content recognition will be a mixture of automated and interactive processes (e.g. "Is that you?"). Associated users would then be informed about new data available about them. With data control level II comes also the ability to prohibit a future linking of this data to a user's profile. However, since it only focusses on linkage, this level does not grant any unpublishing rights to the user. In order to actually unpublish data, we define the next level. To our knowledge there exists no social network application today that provides this kind of functionality.

Data Control Level III. Like level II, but including the possibility to unpublish data about a user from network. In order to provide these features the associated users will have to be provided with rights so that even users who do not own the data are able to enforce their rights upon the data.

Data Control Level IV. Like level III, but including usage control mechanisms that extend the control over user-owned data to the control over usage outside of the infrastructure (e.g. on offline machines etc.). These mechanisms would provide the possibility to constrain the usage of personal data in scenarios like: view, change or copy. This includes the necessity for trusted platforms, monitors, signalers and a rigid policy management. Policies could be generated from configurations, which the data owner applies to his data.

The relation between usage control and data control can be derived from these data control levels. While the first three levels focus on data control of the system's interior, the fourth level considers data that leaves the system onto the platform of a cP in an exterior domain. It is obvious that usage control mechanisms extend the control over personal data; However, only the oP may enforce the usage of his data and no iP will get those exclusive usage control rights. This is no technical short-coming but a social one: If somebody took a picture of the Eiffel Tower, with another person in it, this person is not able to force the photographer to not use this picture for his private purpose or even show or give it to his friends. This fact depicts a fundamental difference to data control levels I-III.

Hence we focus on data control level III leaving the implementation of level IV as optional, in order to establish even more control over personal data. Several usage control architectures have been proposed in literature [1][10][11], which could help to establish data control level IV as an addition to the first three levels.

It is important to acknowledge that our concept of Data Control has to deal with questions of privacy opposed to the right of expression or the right of information, which could be severly limited by our proposed architecture. Although we believe that privacy in an online context is of particular importance and therefor legitimate the use of Data Control tools, it might be necesarry to discuss this matter in future work. However we focus on the conceptional implementation of A DFN and therefor do not particularly discuss moral, ethic or law issues.

3.2 Data Control Scenarios

In this section we describe the basic scenarios to which a data control platform would have to stand up to as well as the main technical components to support these scenarios. After these descriptions, we summarize the conceptional needs in order to implement this data control platform.

Scenario I: Publish Data. The first scenario is in fact the most important one because it deals with the publishing process of data. The analyzing, associating, and categorizing of this data is essential in order to establish data control level III.

We propose a data publish control mechanism. This mechanism, in fact an aggregation of several content recognition and data control mechanisms, comes into action as soon as a user publishes to the network.[3] Before that, the data publish control checks the data for associations to other persons (e.g. face-recognition, names, etc). If the mechanisms find iPs in the data it checks if the iPs are part of the network and in this case notifies the user in question.

If the iP can not be found, because he is (not yet) part of the network, the meta–information will be stored in a database for unknown associations in order to make it possible to link this data at a later date. It is advisable to implement a fine-grained categorization of this meta–information (e.g. gender, eye-color, etc) so the needed resources to search the unknown entries are disburdened. As soon as a new user joins the network his profile would be categorized in the same way and could then be compared to the respective meta–information categories.

The data publish control will also have to deal with different types of data, for example, photos, videos or textual references. While it is conceivable that photo and video recognition algorithms may recognize the simple fact that there is a person in a photo or video, the semantic interpretation of textual data is still impossible. This fact is particularly important when considering the unknown associations: If the algorithm identifies a person is not yet part of the network, then this meta-information can be stored in the database for unknown entries. While it is possible to search textual data for associations to existing users (e.g.

[3] Integration of content recognition mechanisms is desirable but not obligative. A manual check of submitted data through content assignement and a moderator system could replace the automated content recognition, even if this would arise novel questions of practicability.

by name search) it is impossible to find associations to unregistered users. Therefore there will be no unknown associations for textual references. The platforms of newly joined users will have to search all existing textual data (e.g. with their name) to find existing associations to them.

In a DFN, data is first published, and may then be removed by an identified person if she thinks it exposes her privacy. There are various alternatives to schedule the publication of data with respect to its potential later removal. As an extreme, publication may just be immediate, while its removal might take place whenever afterwards requested. This would leave identified persons no time for reaction. As another extreme, publication maybe delayed until all identified persons have agreed. This would hardly be acceptable from the publisher's point of view. As a compromise, one might consider some latency before the actual publishing of the data. In an implementation, the publication control mechanism may adjust the specific latency value depending on the number of identified persons who are currently offline.

If all the published data gets analyzed upon publishing, this means that there will not exist any unchecked data so that given a certain reliability of the content recognition mechanisms all data is associated with its possible iPs. It is easy to see that the whole concept of data control relies heavily on Scenario I and the implied data publish control mechanisms. Correctly implemented, these mechanisms would set the cornerstone for later scenarios like "unpublish" or "control usage".

Scenario II: Find data. There is no need to search the network for associated data that was published after the associated user joined the network: Users may choose to be notified on a push-basis as soon as new associations have been recognized. However, one of the problems of data publish control arises as soon as no existing user name can be found. Consider a defaming commentary about a person who is not yet part of the network. The question now is: How is a newly joined user made aware of associated data that already existed in the network before his arrival. We propose a pull-mechanism here: As soon as a new user arrives, his client could trigger an overview request, and thereby flood the network in search for existing associations (not only the unknown entries but also all textual data). Since the unknown associations are stored by using a fine-grained categorization and processing textual search queries is of relatively low complexity this pull mechanism would not stress the resources in an unaccountable way. Nevertheless, DoS-attacks might be an immanent danger of this process. Therefore this overview request will be only permitted once upon registering, since all later published data will be analyzed by the data publish control.

Scenario III: Confirm Data Association. The third scenario deals with the confirmation of associations to data. If some data is unwanted and the user wants to unpublish, he will first have to confirm his association. Through this confirmation of his association to the data he will also be able to change the privacy settings of the data (e.g. "Only my friends may view this photo").

Obviously this would not prevent malicious users from gaining deletion rights over data that has been falsely connected to them. In order to avoid falsely claimed rights, one could think of a four-eyed confirmation process: the iP himself as well as a third non-involved person would have to testify that the generated associations are true. To make sure that the claimed rights are based on true and trustful associations we therefore suggest the implementation of a "web of trust" [3] in which the authenticy of certificates gets validated through mutual affirmation. Trust levels could be assigned to every user to evaluate the trustworthiness of the user in question [7]. Persons with a high trust level, called notaries, could be consulted to confirm claimed rights of users. In such a network the confirmation of the associated user alone would not be enough to answer the "Is that you"- question .

To get the desired rights, users will have to contact notaries, who would then approve or reject the request. As soon as this confirmation is made the user would get the right to change privacy settings or unpublish the data in question.

The web of trust principles could also be extended to place the discussion about the borderlin between privacy and right of expression in the hand of the users themselves. For example one can think of a system in which the user is not able to directly delete confirmed associated content but has to contact another notary in order to do so. If this notary finds that the content is not offensive he might deny the deletion request. Because of the limited space, we leave this interesting discussion for future work.

Scenario IV: Unpublish Data. Scenario IV uses the same backend described in scenario III. In this backend the user would be able to get an overview of all the associations to him that have been found by the publish control mechanism. Given the correct and trustworthy confirmation of an association to a piece of data in the network (assured during the confirmation process) the user would now be able to hit a "delete"-button and thereby force the publisher's platform to unpublish the data. The implementation of this unpublish process itself should be relatively trivial. By implementing this possibility to unpublish data, data control level III would be established.

Unpublishing the data does not necessarily imply the actual deletion of the data from the network. It may be desired to change the privacy settings of the data in question and thereby manipulate it's visibility or access conditions. The oP will of course have to notified of any change in publish or privacy settings of his data.

3.3 Data Control Requirements

Our main goal is to establish data control level III, thus enabling the user to get an overview over data about him published in the network and the possibility to unpublish. To achieve this goal we formulate the following needs:

- Need for data publish control mechanism that analyzes all published data for associations to existing users through various existing or yet-to-be-developed content recognition mechanisms.

- Need for interactive confirmation process in order to verify that found associations are true.
- Need for processes that confirm claimed associations and make them trustworthy (e.g. web of trust).
- Need to implement a user interface in which he can get an overview of all data associated with him (confirmed or unconfirmed).
- Need to deliver deletion rights to the user for data that has been confirmed to be associated with him.
- Need to categorize unknown associations and store them in a database for later association to newly arrived users.
- Need to couple an automated overview request with the registration process in order to search unknown associations and textual references for connections to the newly arrived user.

4 Solution Proposal

In this section we propose conceptual technical solutions for the data control paradigm. We discuss the pros and cons of two different approaches: a centralized server-client-architecture opposed to a decentralized P2P network. After this discussion we conclude in a general recommendation as well as brief description about necessary next steps.

It is importantion to note, that we intent to publish our architecture under a Open Source License in order to intensify community participation and facilitate integration with other existing approaches.

4.1 Centralized Architecture

It is conceivable to enhance existing server-client architectures such as the *Facebook Open Platform* with data control mechanisms in order to improve the control over personal data in the network. Therefore our proposal in this section is to use the Facebook Open platform and extend it with those mechanisms to provide additional functionality. In a centralized architecture all data is stored on a single server which is usually managed by a service provider. In order to achieve data control levels in an architecture of this kind the following measures will have to be taken:

Data Control Backend. A user interface has to be implemented in which users are provided with an overview of their own data, data about them and possible associations to data. This backend will have to provide the functionality to modify the privacy settings of data belonging to a user or about a user as well as unpublishing of the content in question. Also it will serve as a notification board for newly recognized associations. From this backend the user will be able to confirm associations and thereby claim his rights to modify the privacy settings of this content or completely unpublish it.

Server-Side Data Publish Control. Each piece of data submitted to the network will have to pass through a server-side data publish control before being published. This data publish control consists of several mechanisms, including the content-recognition algorithms as well as categorization of unknown content. Also this data publish control will deal with the associations to existing users and notify them of possible references. Unknown associations (for example to persons who are not yet part of the network) will be stored in a database for later processing.

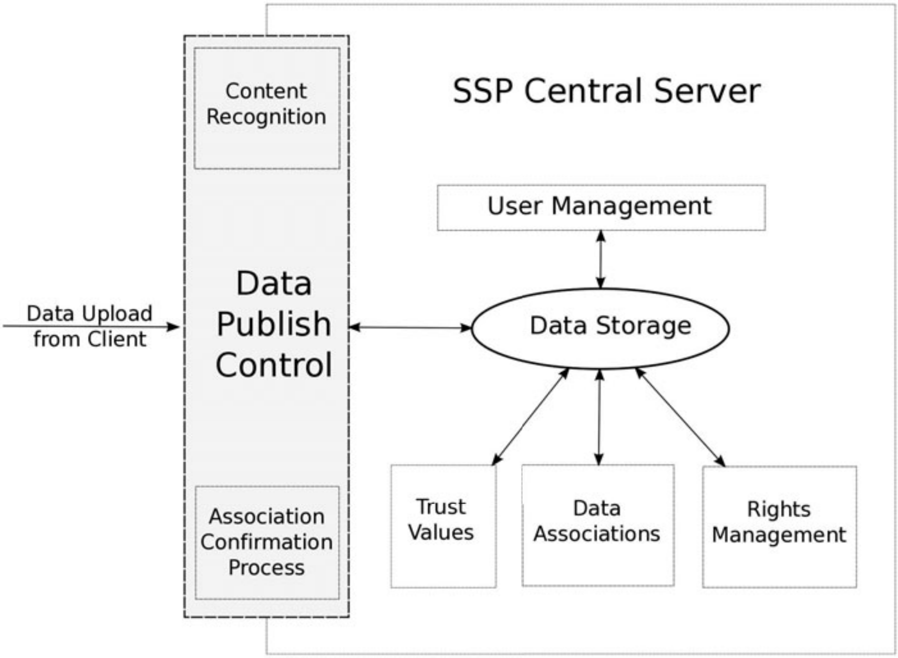

Fig. 2. Structural Components of a centralized DDFN

Confirming associations through web of trust. In order to confirm and verify the associations that have been found by the data publish control, it is necessary to implement a web of trust[4] in which users are assigned to trust levels. Users with high trust levels, so called notaries, will be required to verify associations to pieces of data and thereby enable users to claim their rights upon content that does not belong to them but references them in some way.

Search the network after registration. To make newly arrived users aware of pre-existing data about them, that has been published before they joined the network, the registration process has to be modified. Firstly, the user would have to provide the platform with information (e.g. with fotos, names, etc) that can be

[4] One can think of alternatives, however the principle of a web of trust fits the purpose very well here.

used to search the network for similar data entries. Secondly, this information will be used just after registration in order to search the unknown entries database and all textual references in the network. This search request is only triggered once upon registering, since afterward all newly published data will be recognized by the data publish control.

Usage Controlled Clients. If it is desired to establish data control level IV all clients will have to be equipped with a usage control platform. For example, Zhang et al.'s approach consisting of a SELinux as a trusted platform combined with mandatory access control policies (MAC-policies), would ensure that data belonging to users would not be used in an undesired way as soon as it leaves the network.

5 Decentralized P2P Architecture

The popular term "P2P" describes a decentralized infrastructure in which peers directly and autonomously exchange data with other peers of their choice. In a P2P network there is no single service provider and no central server on which all data is stored. To the contrary all information is stored on the peers themselves and made available from there.

In the past this structure has been particularly successful in file sharing domains. One of the most popular representatives of the P2P file sharing platforms is *Gnutella*[5], a specific network protocol on which many popular file sharing clients (e.g. *LimeWire*)[6] are based on.

However in the last years, research has also considered P2P infrastructures for various other fields of application such as consumer-to-consumer (C2C) infrastructures for platforms like eBay or Amazon [6]. *Tribler*[7] uses a P2P infrastructure for video-on-demand services. Recently there have also been several approaches considering a decentralized P2P infrastructure as the base for a privacy-enabling social network.

A promising approach is *Safebook* by Cutillo et al. [5], in which a three component architecture is proposed: a trusted service for identification, a P2P infrastructure and a setting of matryoshkas, which are trust rings of friends (peers) surrounding a user on which the user may then store of publish data.

We propose an extension of *Safebook* by adding functionalities, which establish data control over personal data. In order to do so we need to address the following issues:

Data publish controlled peers. In a P2P network there is no central computing instance, which may detect data associations to other users. Hence, the data publish control, including the content recognition mechanisms must be applied on a local client software located on the peer itself. Newly published data will be checked for associations to friends of the publisher. If the data identifies

[5] `http://rfc-gnutella.sourceforge.net/` Last checked: 25.03.10.

[6] `http://www.limewire.com/` Last checked: 25.03.10.

[7] `http://www.tribler.org/` Last checked: 25.03.10.

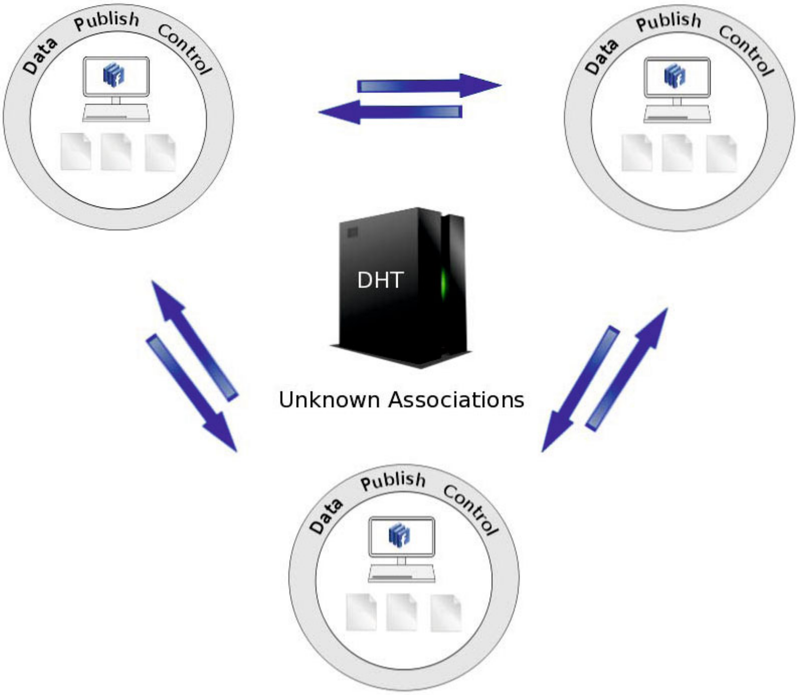

Fig. 3. Conceptional sketch of data control in a P2P network

some of the publisher's friends, these friends will be instantly notified of the new association and may then claim rights of this data.

However it might be possible that the submitted data holds associations to users that are not directly connected to the publisher herself. In a privacy-enabling P2P environment there is no way to check for associations of all users in the network instantly. Only those peers directly or to some small degree connected to the own peer are known. Since there is no central storage device holding a list of all users in the network, these associations will have to be considered unknown and will therefore be treated the same as truly unknown associations to users who are actually not registered in the network as long as no iP has been found.

The unknown associations have to be made available for later processing both for users in the network who are not directly connected to one's peer as well as for newly arrived users. Therefore the metadata collected out of the content recognition mechanisms will be stored in a distributed hash table (DHT). In order to preserve privacy aspects of the network this DHT will only hold metadata with links to the data itself. No personal data will be made available in the DHT.

Search the DHT after registration. As soon as a new peer enters the network its user might want to know if there is already data about him spread in the network.

In order to find out, he will need to provide basic information about himself to his client during the registering process, such as name, biometrical data from photos etc. After the registration the client will then trigger an overview request.

Search the DHT continuously. In contrast to a centralized approach, the finding of new associations in a P2P environment, will not be a passive action solely. On the one hand, associations are collected passively, considering data that has been published by users that are connected to a user's peer (e.g. friends, friends-of-friends). However data that has been published by users that are not known to the user's peer may also contain associations to him. The metadata of these unknown associations is stored in a DHT, which will be searched continuously, and incrementally by the user's client itself. This search process might take a while for the first time; however, as soon as the DHT is searched completely the search algorithm may concentrate on updates and will therefore guarantee a timely reaction to newly published data with associations to oneself.

Confirming associations through web of trust. Similarly to a centralized data control approach each peer will have to be assigned with trust levels. In order to claim rights over associated data, the user will have to contact a highly trusted individual (notary) who may then accept or decline the right claim. After an acceptance by a notary, the user may send out requests to the publisher's peer in order to change privacy settings or unpublish the data in question.

External publish control on local clients. As soon as the user successfully claimed rights over data associated to him, he might want to change the privacy settings of the data in question or even be able to unpublish it. This task is not trivial, since the data and the privacy settings are not published to an external server control by a SSP. On the contrary the data will be located on the peer itself. In order to externally trigger the change of these local settings the client platform will have to be able to enforce external right applications from other users.

6 Centralized vs. Decentralized

It is imaginable to enhance existing server-client architectures such as the *Facebook Open Platform* with data control mechanisms in order to improve the control over personal data in the network. For example one could use the *Facebook Open platform* and extend it with those mechanisms and additional functionality.

Opposed to the advantages of central storage of data there are serious concerns of privacy due to the power of control of the single service provider. Even if most of the security-disadvantages of a centralized system such as intrusion attempts, DoS-Attacks, etc. could be handled, the single service provider (SSP) would still be able to use the accumulated data according their own discretion. Similarly to *facebook*, the SSP could sell data to parties outside of the network and would therefore undermine our described notion of data control. Even terms of usage, which forbid the sale of this information, would be no guarantee for the user's control over his data, since the SSP might change his terms of usage at any time.

Because of these main problems of any centralized architecture we conclude that a data controlled network established in a centralized manner would raise serious scalability, resource and privacy problems and is therefore not first choice in order to implement a DFN.

A decentralized P2P solution with a special focus on privacy (such as *Safebook*) promises solutions for the shortcomings of a centralized approach. Nevertheless further research has to take place concerning P2P-inherent questions and problems, such as scalability, availability and practicability (e.g. of search algorithms in a distributed network). However, these questions are orthogonal to our research and therefore not in our center of attention.

7 Conclusion

We propose the new paradigm of user-centered data control in order to address the rising challenges for privacy and security in social networks. This concept is opposed to traditional content-centric networks where users are not able to enforce their rights on personal data once it has been published.

We identify different data control levels reaching from simple linkage control as in existing online social networks to the possibility to actually unpublish unwanted data no matter who published it. Additionally, usage control mechanisms could be applied in order to extend this control to exterior domains after the data left the network domain.

Basic usage scenarios such as “publish data”, “find data” or “unpublish data” are described in this paper through which basic technological challenges and necessary components are outlined.

We come to the conclusion that a defamation-free network (DFN), in which users have control over private data, no matter if published by them or by others, is realizable both in a centralized or decentralized manner. However centralized architectures come with serious privacy issues concerning the information sovereignty of the singel service provider. Therefore we suggest a P2P network architecture for our DFN and point out the main technological challenges and requirements.

8 Future Work

Following this work it is necessary to explore the possibility of implementing data control functions in a distributed environment such as *Safebook* [5]. It is imperative to provide a prototypical architecture based on one of the P2P network approaches. We will especially have to deal with questions of network scalability and availability of the data, with respect to the metadata as well as the data itself. In order to make the metadata available at all times distributed servers could be provided by some of the peers themselves. Questions about motivation and incentives for such a behavior arise and should be addressed. Tests, evaluations and simulations of this architecture have to be included as well as recommendations for the extension of the prototype. Formal and theoretical aspects are of high importance especially when considering the data control rights management as well as the verification of data control security intentions. Therefore, a formal specification of Data Control and it's functions in a defamation-free network are currently under development by the authors.

References

1. Alam, M., Seifert, J.-P., Li, Q., Zhang, X.: Usage control platformization via trustworthy selinux. In: ASIACCS, pp. 245–248 (2008)
2. Buchegger, S., Schiöberg, D., Vu, L.H., Datta, A.: PeerSoN: P2P Social Networking - Early Experiences and Insights. In: Second ACM Workshop on Social Network Systems Social Network Systems 2009, Nürnberg, Germany (March 31, 2009)
3. Caronni, G.: Walking the web of trust. In: WETICE 2000: Proceedings of the 9th IEEE International Workshops on Enabling Technologies, pp. 153–158. IEEE Computer Society, Washington, DC, USA (2000)
4. Castelluccia, C., Kaafar, M.A.: Owner-centric networking (ocn): Toward a data pollution-free internet. In: SAINT 2009: Proceedings of the 2009 Ninth Annual International Symposium on Applications and the Internet, pp. 169–172. IEEE Computer Society, Washington, DC, USA (2009)
5. Cutillo, L.A., Molva, R., Strufe, T.: Safebook: a privacy preserving online social network leveraging on real-life trust. To appear in IEEE Communications Magazine, Consumer Communications and Networking Series (December 2009)
6. Datta, A., Hauswirth, M., Aberer, K.: Beyond web of trust: Enabling p2p e-commerce. In: CEC, pp. 303–312 (2003)
7. Eilers, F., Nestmann, U.: Deriving trust from experience. In: Degano, P., Guttman, J.D. (eds.) FAST 2009. LNCS, vol. 5983, pp. 36–50. Springer, Heidelberg (2010)
8. Hafner, M., Breu, R.: Security Engineering for Service-Oriented Architectures. Springer, Heidelberg (2009)
9. Park, J., Sandhu, R.S.: The ucon$_{abc}$ usage control model. ACM Trans. Inf. Syst. Secur. 7(1), 128–174 (2004)
10. Pretschner, A.: An overview of distributed usage control. In: Proc. of KEPT 2009 International Conference, pp. 25–33 (July 2009)
11. Pretschner, A., Büchler, M., Harvan, M., Schaefer, C., Walter, T.: Usage control enforcement with data flow tracking for x11. In: 5th International Workshop on Security and Trust Management, STM 2009 (2009)