Resolving Multi-party Privacy Conflicts in Social Media

Jose M. Such, *Member, IEEE,* Natalia Criado

Abstract—Items shared through Social Media may affect more than one user's privacy — e.g., photos that depict multiple users, comments that mention multiple users, events in which multiple users are invited, etc. The lack of multi-party privacy management support in current mainstream Social Media infrastructures makes users unable to appropriately control to whom these items are actually shared or not. Computational mechanisms that are able to merge the privacy preferences of multiple users into a single policy for an item can help solve this problem. However, merging multiple users' privacy preferences is not an easy task, because privacy preferences may conflict, so methods to resolve conflicts are needed. Moreover, these methods need to consider how users' would actually reach an agreement about a solution to the conflict in order to propose solutions that can be acceptable by all of the users affected by the item to be shared. Current approaches are either too demanding or only consider fixed ways of aggregating privacy preferences. In this paper, we propose the first computational mechanism to resolve conflicts for multi-party privacy management in Social Media that is able to adapt to different situations by modelling the concessions that users make to reach a solution to the conflicts. We also present results of a user study in which our proposed mechanism outperformed other existing approaches in terms of how many times each approach matched users' behaviour.

Index Terms—Social Media, Privacy, Conflicts, Multi-party Privacy, Social Networking Services, Online Social Networks

✦

1 INTRODUCTION

HUNDREDS of billions of items that are uploaded to Social Media are co-owned by multiple users [1], yet only the user that uploads the item is allowed to set its privacy settings (i.e., who can access the item). This is a massive and serious problem as users' privacy preferences for co-owned items usually conflict, so applying the preferences of only one party risks such items being shared with undesired recipients, which can lead to privacy violations with severe consequences (e.g., users losing their jobs, being cyberstalked, etc.) [2]. Examples of items include photos that depict multiple people, comments that mention multiple users, events in which multiple users are invited, etc. Multi-party privacy management is, therefore, of crucial importance for users to appropriately preserve their privacy in Social Media. A solution for this could be to merge individual privacy preferences into a joint sharing policy for the item. This joint policy would be valid as long as individual preferences do not conflict. However, there are many situations in which privacy preferences may conflict. For instance, imagine that Alice and Bob are depicted together in a photo in which Bob appears drunk, and that Alice would like to share the photo with her friend Charlie, but Bob would not like to share the photo with Charlie because Bob feels embarrassed about his looks

- *Jose M. Such is with the School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA, UK.*
  e-mail: j.such@lancaster.ac.uk
- *Natalia Criado is with the School of Computing and Maths, Liverpool John Moores University, Byrom Street, Liverpool L3 3AJ, UK.*
  e-mail: n.criado@ljmu.ac.uk

in the photo and Charlie is just a distant acquaintance of him.

There is recent evidence that users very often negotiate collaboratively to achieve an agreement on privacy settings for co-owned information in Social Media [3], [4]. In particular, users are known to be generally open to accommodate other users' preferences, and they are willing to make some concessions to reach an agreement depending on the specific situation [4]. However, current Social Media privacy controls solve this kind of situations by only applying the sharing preferences of the party that uploads the item, so users are forced to negotiate *manually* using other means such as e-mail, SMSs, phone calls, etc. [5] — e.g., Alice and Bob may exchange some e-mails to discuss whether or not they actually share their photo with Charlie. The problem with this is that negotiating *manually* all the conflicts that appear in the everyday life may be time-consuming because of the high number of possible shared items and the high number of possible accessors (or targets) to be considered by users [2]; e.g., a single average user in Facebook has more than 140 friends and uploads more than 22 photos [6].

Computational mechanisms that can automate the negotiation process have been identified as one of the biggest gaps in privacy management in social media [3], [4], [5], [7], [8]. Note that the aim of these mechanisms would not be that users lose control by providing a *fully* automated solution, as this has been proven not to achieve satisfactory results when it comes to privacy management in social media [9]. Instead, these mechanisms would *suggest* a possible solution to the conflicts that users would need to accept to be finally applied — if

users do not accept the suggestion they will need to enter into a manual negotiation by other means. Therefore, the *main challenge* is to propose solutions that can be accepted most of the time by all the users involved in an item (e.g., all users depicted in a photo), so that users are forced to negotiate manually as little as possible, thus minimising the burden on the user to resolve multi-party privacy conflicts.

Very recent related literature proposed mechanisms to resolve multi-party privacy conflicts in social media [2], [10], [11], [12], [13], [14]. Some of them [10], [11] need too much human intervention during the conflict resolution process, by requiring users to solve the conflicts *manually* or close to *manually*; e.g., participating in difficult-to-comprehend auctions for each and every co-owned item. Other approaches to resolve multi-party privacy conflicts are more automated [2], [12], [13], but they only consider one fixed way of aggregating user's privacy preferences (e.g., veto voting [2]) without considering how users would actually achieve compromise and the concessions they might be willing to make to achieve it depending on the specific situation. Only [14] considers more than one way of aggregating users' privacy preferences, but the user that uploads the item chooses the aggregation method to be applied, which becomes a unilateral decision without considering the preferences of the others. All of this makes the above mechanisms unable to adapt to different situations that may motivate different users' concessions, suggesting solutions that may not be acceptable by all users, so that users may need to end up negotiating manually most of the time.

In this paper, we present the first computational mechanism for social media that, given the individual privacy preferences of each user involved in an item, is able to find and resolve conflicts by applying a different conflict resolution method based on the concessions users' may be willing to make in different situations. We also present a user study comparing our computational mechanism of conflict resolution and other previous approaches to what users would do themselves manually in a number of situations. The results obtained suggest our proposed mechanism significantly outperformed other previously proposed approaches in terms of the number of times it matched participants' behaviour in the study.

The remainder of the paper is organised as follows. Section 2 provides some background on individual privacy preferences and the problem statement. Section 3 provides an overview of our proposed computational mechanism for multi-party privacy conflicts resolution. Section 4 describes the mechanism to detect multi-party privacy conflicts by analysing users' individual privacy preferences. Section 5 presents the method to estimate concessions and, based on this, suggest solutions to multi-party privacy conflicts. Section 6 details the method, participants and results of the user study we conducted to compare our proposal with other state-of-the-art approaches. Section 7 discusses related work. Finally, Section 8 contains some concluding remarks.

2 BACKGROUND

Assume a finite set of users U, where a finite subset of *negotiating* users $N \subseteq U$, negotiate whether they should grant a finite subset of *target* users[1] $T \subseteq U$ access to a particular co-owned item. For instance, Alice and Bob (negotiating users) negotiate about whether they should grant Charlie (target user) access to a photo of them depicted together. For simplicity and without loss of generality, we will consider a negotiation for one item over the course of this paper — e.g., a photo that depicts the negotiating users together — and hence, we do not include any additional notation for the item in question.

2.1 Individual Privacy Preferences

Negotiating users have their own individual privacy preferences about the item — i.e., to whom of their online friends they would like to share the item if they were to decide it unilaterally. In this paper, we assume negotiating users specify their individual privacy preferences using group-based access control, which is nowadays mainstream in Social Media (e.g., Facebook lists or Google+ circles), to highlight the practical applicability of our proposed approach. Note, however, that other access control approaches for Social Media could also be used in conjunction with our proposed mechanism — such as relationship-based access control [15], [16], [17], or (semi-)automated approaches like [18], [19], [20]. To illustrate this, Appendix A describes how relationship-based access control could be used in conjunction with our mechanism. Note also that our approach does not necessarily need users to specify their individual privacy preferences for each and every item separately, they could also specify the same preferences for collections or categories of items for convenience according to the access control model being used —e.g., Facebook users can specify the same preferences for all the photos in a photo album.

Mainstream Social Media (Facebook, Google+, etc.) have predefined groups and also allow users to define their own groups, each of which is composed of a set of friends. Access to items (photos, etc.) can be granted/denied to groups, individuals or both (e.g., all Friends have access to a photo except Charlie). We formally define a group $G \subseteq U$ as a set of users, and the set of all groups defined by a particular user u as $\mathcal{G}_u = \{G_1, \ldots, G_l\}$, so that $\bigcap_{G \in \mathcal{G}_u} G = \emptyset$. For instance, Alice may have defined the following groups $\mathcal{G}_{Alice} = \{\text{CloseFriends}, \text{Family}, \text{Coworkers}\}^2$ to organise her online friends.

---

1. Note that we defined the set of target users as a subset of the users; i.e., without forcing it to satisfy a particular property, to remain as general as possible. However, the set of target users could be further qualified as a particular subset of users satisfying any property without changing the subsequent formalisation; e.g., the set of target users could be defined as the union of all of the negotiating users' online friends.

2. For simplicity, we represent groups using a meaningful label

Definition 1: A privacy policy P is a tuple $P = \langle A, E \rangle$, where A is the set of groups granted access and $E \subseteq U$ is a set of individual user exceptions.

According to most Social Media infrastructures, the semantics of a group-based privacy policy are as follows: $P.A$ are the groups that are *authorised* (or granted) access to the item; i.e., all users belonging to these groups are granted access to the item; and $P.E$ are a set of individual exceptions; i.e., either users in the authorised groups who are denied access individually, or users who are granted access individually because they are in the unauthorised groups —which are the groups not explicitly granted access. Continuing the example above, Alice now defines her individual privacy policy for an item as $P_{Alice} = \langle \{\texttt{CloseFriends}\}, \{\texttt{Charlie}\} \rangle$, which means that Alice wants to share the item only with CloseFriends but excluding Charlie.

2.2 Problem Statement

The problem we deal with is the following: *how the set of negotiating users $N = \{n_1, \ldots, n_k\}$ who co-own an item but that have different individual (and possibly conflicting) privacy policies $P_{n_1}, \ldots, P_{n_k}$ for that item can agree on to whom, from the set of the target users $T = \{t_1, \ldots, t_m\}$, the item should be shared with?* This problem can be decomposed into:

1) Given the set of individual privacy policies $P_{n_1}, \ldots, P_{n_k}$ of each negotiating user for the item, how can we identify if at least two policies have contradictory decisions — or *conflicts* — about whether or not granting target users T access to the item[3].

2) If conflicts are detected, how can we propose a solution to the conflicts found that respects as much as possible the preferences of negotiating users N.

3 MECHANISM OVERVIEW

We propose the use of a mediator that detects conflicts and suggests a possible solution to them. For instance, in most Social Media infrastructures, such as Facebook, Twitter, Google+ and the like, this mediator could be integrated as the back-end of Social Media privacy controls' interface; or it could be implemented as a Social Media application — such as a Facebook app — that works as an interface to the privacy controls of the underlying Social Media infrastructure. In a nutshell, the process the mediator follows is:

1) The mediator inspects the individual privacy policies of all users for the item and flags all the

conflicts found (as described in Section 4). Basically, it looks at whether individual privacy policies suggest contradictory access control decisions for the same target user.

2) The mediator proposes a solution for each conflict found. To this aim, the mediator estimates (as described in Section 5) how willing each negotiating user may be to concede by considering: her individual privacy preferences, how sensitive the particular item is for her, and the relative importance of the conflicting target users for her.

If all users accept the solution proposed, it will be applied. Otherwise, users will need to turn into a *manual* negotiation by other means.

4 CONFLICT DETECTION

We need a way to compare the individual privacy preferences of each negotiating user in order to detect conflicts among them. However, each user is likely to have defined different groups of users, so privacy policies from different users may not be directly comparable. To compare privacy policies from different negotiating users for the same item, we consider the *effects* that each particular privacy policy has on the set of target users T. Privacy policies dictate a particular action to be performed when a user in T tries to access the item. In particular, we assume that the available actions are either 0 (denying access) or 1 (granting access). The action to perform according to a given privacy policy is determined as follows[4]:

Definition 2: Given an user $n \in N$, her groups $\mathcal{G}_n$, her individual privacy policy $P_n = \langle A, E \rangle$, and a user $t \in T$; we define the action function as:

$$act(P_n, t) = \begin{cases} 1 & \text{if } \exists G \in \mathcal{G}_n : t \in G \wedge G \in P_n.A \wedge t \notin P_n.E \\ 1 & \text{if } \exists G \in \mathcal{G}_n : t \in G \wedge G \notin P_n.A \wedge t \in P_n.E \\ 0 & \text{otherwise} \end{cases}$$

We also consider so-called *action vectors* $\vec{v} \in \{0, 1\}^{|T|}$; i.e., complete assignments of actions to all users in T, such that $v[t]$ denotes the action for user $t \in T$. When a privacy policy is applied to the set of users T, it produces such an action vector:

Definition 3: The *action vector* produced by privacy policy P when applied to T is $\vec{v} = (v_1, \ldots, v_{|T|})$, where $v[t] = act(P, t)$.

Given the action vectors produced by the privacy policies of all negotiating users, we are now in a position to be able to detect whether there are any conflicting actions suggested for the same target user. That is, if all the action vectors assign the same action for all target users, then there is no conflict. Otherwise, there are at

---

3. We focus on detecting conflicts once we know all the parties that co-own an item and have their individual privacy policies for the item. We are not proposing a method to automatically detect which items are co-owned and by whom they are co-owned. This is a different problem that is out of the scope of this paper. Refer to [21] as an example of a method to detect the people that is depicted in a photo with very high accuracy for Social Media.

4. Note that the definition of this function will vary according to the access control model used, but it will be defined in a similar way. That is, the idea is to be able to know, given a target user t, whether the privacy policy will grant/deny t access to the item regardless of the access control model being used. For instance, Appendix A describes this function for relationship-based access control.

least two action vectors that assign different actions to the same target user, and there is a conflict. In other words, a conflict arises when some negotiating users would like to grant access to one target user while the others would not. Formally:

Definition 4 (conflict): Given a set of negotiating users N and a set of target users T; a target user $t \in T$ is said to be *in conflict* iff $\exists a, b \in N$ with individual privacy policies P_a and P_b respectively, so that $v_a[t] \neq v_b[t]$.

Further, we say that the set of *users in conflict* $C \subseteq T$, is the set that contains all the target users that are in conflict.

Algorithm 1 Conflict Detection

Input: N, $P_{n_1}, \ldots, P_{n_{|N|}}$, T
Output: C
1: **for all** $n \in N$ **do**
2:     $\vec{v}_n \leftarrow$ Apply P_n to T
3: **end for**
4: $C \leftarrow \emptyset$
5: **for all** $t \in T$ **do**
6:     **if** $\exists a, b \in N, v_a[t] \neq v_b[t]$ **then**
7:         $C \leftarrow C \cup \{t\}$
8:     **end if**
9: **end for**

The mediator runs Algorithm 1 to detect conflicts by harvesting the *users in conflict* set C. The complexity of the algorithm is polynomial and it depends on the number of negotiating users and target users $\mathcal{O}(|N|^2 \times |T|)$. If Algorithm 1 does not detect any conflict — i.e., $C = \emptyset$, it will return to the users without changes to their preferred privacy policies. If Algorithm 1 detects conflicts, the mediator will then run the conflict resolution module, which is described in the following section.

Example 1: Assume a set of users $U = \{\texttt{Alice}, \texttt{Bob}, \texttt{Charlie}, \texttt{Dan}, \texttt{Eve}, \texttt{Frank}\}$. Negotiating users $N = \{\texttt{Alice}, \texttt{Bob}\}$ are in the process of deciding to which target users $T = \{\texttt{Charlie}, \texttt{Dan}, \texttt{Eve}, \texttt{Frank}\}$ they grant access to a photo in which both of them are depicted. Negotiating users defined the following groups: Alice defined $\mathcal{G}_{\texttt{Alice}} = \{\texttt{MyFriends}\}$ so that $\texttt{MyFriends} = \{\texttt{Charlie}, \texttt{Dan}, \texttt{Eve}\}$; and Bob defined $\mathcal{G}_{\texttt{Bob}} = \{\texttt{CloseFriends}, \texttt{Family}\}$ so that $\texttt{CloseFriends} = \{\texttt{Charlie}, \texttt{Eve}\}$ and $\texttt{Family} = \{\texttt{Dan}, \texttt{Frank}\}$. Now, assume that negotiating users have the following individual privacy policies for the photo: Alice has $P_{\texttt{Alice}} = \langle\{\texttt{MyFriends}\}, \{\texttt{Eve}\}\rangle$ so that $\vec{v}_{\texttt{Alice}} = (1, 1, 0, 0)$ — i.e., only Charlie and Dan would be granted access to the photo; and Bob has $P_{\texttt{Bob}} = \langle\{\texttt{CloseFriends}, \texttt{Family}\}, \emptyset\rangle$ so that $\vec{v}_{\texttt{Bob}} = (1, 1, 1, 1)$ — i.e., all target users Charlie, Dan, Eve and Frank would be granted access to the photo. As $v_{\texttt{Alice}}[\texttt{Eve}] \neq v_{\texttt{Bob}}[\texttt{Eve}]$ and $v_{\texttt{Alice}}[\texttt{Frank}] \neq v_{\texttt{Bob}}[\texttt{Frank}]$, the set of users in conflict is $C = \{\texttt{Eve}, \texttt{Frank}\}$.

5 CONFLICT RESOLUTION

After conflicts are detected, the mediator suggests the best possible solution according to negotiators' preferences. To this aim, the mediator models negotiation behaviour; i.e., to what extent negotiating users might be willing to concede in favour of the actions proposed by others. We shall firstly explain the method by which the mediator estimates the willingness of each user to accept changing their preferred action for each conflict detected. Secondly, we shall explain how the mediator estimates whether each negotiating user would actually concede or not for a given conflict based on her willingness to change her preferred action for the conflict and on the implications for the other negotiating users — i.e., how beneficial/detrimental this change would be for her and the others. Finally, we shall describe how the mediator computes a solution to the conflicts based on the concessions users would do.

5.1 Estimating the Willingness to change an action

In order to find a solution to the conflict that can be acceptable by all negotiating users, it is key to account for how important is for each negotiating user to grant/deny access to the conflicting target user. In particular, the mediator estimates how *willing* a user would be to change the action (granting/denying) she prefers for a target agent in order to solve the conflict based on two main factors: the sensitivity of the item and the relative importance of the conflicting target user.

5.1.1 Estimating Item Sensitivity

If a user feels that an item is very sensitive for her[5], she will be less willing to accept sharing it than if the item is not sensitive for her [22], [23]. One way of eliciting item sensitivity would be to ask the user directly, but this would increase the burden on the user. Instead, the mediator estimates how sensitive an item is for a user based on how strict is her individual privacy policy for the item [19], so that the stricter the privacy policy for the item the more sensitive it will be. Intuitively, the lower the number of friends granted access, the stricter the privacy policy, hence, the more sensitive the item is. Moreover, not all friends are the same; i.e., users may feel closer to some friends than others and friends may be in different groups representing different social contexts. Thus, both the group and the strength of each relationship are considered when estimating the strictness of privacy policies and, therefore, the sensitivity of items.

The mediator can use any of the existing tools to automatically obtain relationship strength (or *tie strength*) values for all the user's friends for particular Social Media infrastructures such as Facebook [24], [25] and Twitter [26] with minimal user intervention. Even if the mediator would not be able to use these tools, users

---

5. Note that we particularly stress that an item is sensitive for someone. This is because the same item may be seen as having different sensitivity by different people.

could be asked to self-report their tie strength to their friends, which would obviously mean more burden on the users but would still be possible. Whatever the procedure being used, the mediator just assumes that the tie strength value assigned for each pair of friends a and b is given by a function $\tau(a, b)$, so that $\tau : U \times U \to \{0, \ldots, \delta\}$, where δ is the maximum positive integer value in the tie strength scale used[6].

Based on these values, the mediator considers how strict is a user's individual privacy policy as an *estimate of* the sensitivity of an item by calculating the minimum tie strength needed in each group to have access to the item and averaging it across groups. That is, if a privacy policy only grants users with close relationships (i.e., friends with high tie strength values) access to an item, then the item will be estimated as sensitive, since the privacy policy is very strict (i.e., the average minimum tie strength across groups to have access to the item is very high). On the contrary, if a privacy policy grants users with low tie strengths across groups, then the item will be estimated as less sensitive, since the privacy policy is less strict.

Definition 5: Given a user $n \in N$, her groups $\mathcal{G}_n$, and her individual privacy policy P_n for an item, the sensitivity of the item for n is estimated as:

$$\mathcal{S}_n = \frac{1}{\mid \mathcal{G}_n \mid} \sum_{G \in \mathcal{G}_n} \mathcal{T}_n(G)$$

where $\mathcal{T}_n(G)$ is the strictness of the privacy policy in group G, defined as the minimum tie strength needed in group G to have access to the item:

$$\mathcal{T}_n(G) = \min_{t \in G} f(n, t)$$

and $f(n, t)$ is based on the tie strength between users n and t. However, this function considers differently situations where t is given access and situations where t is denied access. In particular, if user t is granted access, then function f returns the tie strength between users n and t. On the contrary, if user t is denied access, then this user must not be considered when determining the policy strictness for the group and function f returns the maximum tie strength value (recall that $\mathcal{T}_n(G)$ is defined as the minimum value returned by function f for all users in a group). More formally, $f(n, t)$ is defined as follows:

$$f(n, t) = \begin{cases} \tau(n, t) & \text{iff } act(P_n, t) = 1 \\ \delta & \text{iff } act(P_n, t) = 0 \end{cases}$$

5.1.2 Estimating the relative importance of the conflict

Now the focus is on the particular conflicting target user — i.e., the target user for which different negotiating

---

6. The maximum tie strength value δ depends on the tool used. For example, in Fogués et al. [24] $\delta = 5$; i.e., six levels of tie strength, which would map to, for instance, the friend relationship as: 0-no relationship, 1-acquaintance, 2-distant friend, 3-friend, 4-close friend, 5-best friend.

users would prefer a different action (denying/granting access to the item). The mediator estimates how important a conflicting target user is for a negotiating user by considering both tie strength with the conflicting target user and the group the conflicting target user belongs to, which are known to play a crucial role when deciding who can see which items — refer to [27], [28], [29] and [18], [20], [30] respectively. For instance, Alice may decide she does not want to share a party photo (in which she appears with her class mates) with her mother, who has a very close relationship to Alice (i.e., tie strength between Alice and her mother is high). If some of her class mates want to make the party photo public to anyone (which entails that Alice's mother will be able to see the photo), it is plausible to believe that Alice would be really unwilling to share the party photo with her mother (i.e., not sharing the photo with her mother is very important to Alice). Another example would be a photo in which Alice is depicted together with some friends with a view to a monument, and Alice wants to share it with all her friends. If some of her friends that appear in the monument photo want to share it with Alice's acquaintances, it is plausible to believe that Alice may be willing to share the monument photo with acquaintances too as she is willing to share it with all her friends (not sharing the photo with acquaintances may not be very important to Alice). Thus, the mediator estimates the relative importance of a particular conflicting user considering both the tie strength with this user in general and within the particular group she belongs to. In particular, the mediator estimates the relative importance a conflicting target user has for a negotiating user as the difference between the tie strength with the conflicting user and the strictness of the policy for the group the conflicting user belongs to. If the conflicting target user does not belong to any group of the negotiator; then the relative importance is estimated considering the item sensitivity instead as there is no group information.

Definition 6: Given a user $n \in N$, her groups $\mathcal{G}_n$, and a conflicting user $c \in C$, the relative importance of c for n is estimated as follows:

$$I_n(c) = \begin{cases} \mid \mathcal{T}_n(G) - \tau(n, c) \mid & \text{if } \exists G \in \mathcal{G}_n : c \in G \\ \mid \mathcal{S}_n - \tau(n, c) \mid & \text{otherwise} \end{cases}$$

For instance, assume Alice would like to share with all her friends —i.e., $\mathcal{T}_{Alice}(Friends) = 1$— but not with Charlie, who is close friend of her —i.e., $\tau(Alice, Charlie) = 5$. The relative importance would be calculated as $I_{Alice}(Charlie) = \mid 1 - 5 \mid = 4$, which means that the action Alice prefers for Charlie is quite important to her; e.g., Alice could be creating an event in which she invites all her friends except Charlie because the event is a surprise for Charlie's birthday, so sharing with Charlie would mean ruining the surprise party. In the very same way, if Alice would like to share an item only with her best friend —i.e., $\mathcal{T}_{Alice}(Friends) = 5$, the relative im-

portance of denying access to an acquaintance would be high too —i.e., if Peter is an acquaintance of Alice such that $\tau(Alice, Peter) = 1$, then $I_{Alice}(Peter) =\mid 5 - 1 \mid= 4$.

5.1.3 Estimating Willingness

Finally, the mediator estimates the *willingness* to change the preferred action (granting/denying) for a conflicting target user accounting for both the sensitivity of the item and the relative importance of the conflicting target user as detailed above. If both sensitivity and relative importance are the highest possible, then the willingness to change should be minimal. On the contrary, if both sensitivity and relative importance are the lowest possible, then the willingness to change should be maximal. Thus, we define willingness as a distance (in a 2-dimensional space) between the values of both item sensitivity and relative importance and the maximum possible values for both — as shown above, both measures are defined in tie strength units and have δ as their maximum value[7]. We chose for this the Canberra distance[8] instead of other distances like Euclidean, Manhattan, or Chebyshev because it is a relative and not absolute distance metric — so that it would work in the same way regardless of the δ value being used.

Definition 7: Given user $n \in N$, her preferred privacy policy P_n, the maximum tie strength value δ, a conflicting target user $c \in C$, the willingness of user n to accept changing her most preferred action for c is a function $\mathcal{W} : N \times C \to [0, 1]$ so that:
$$\mathcal{W}(n, c) = \frac{1}{2} \cdot \left(\frac{\mid \delta - I_n(c) \mid}{\delta + I_n(c)} + \frac{\mid \delta - S_n \mid}{\delta + S_n} \right)$$

Note that the only difference from a 2-dimensional Canberra distance is that we divide by 2 the final result to normalise the willingness into a real value within the [0,1] interval for convenience to model concessions as shown in the following section (Section 5.2).

Example 2: Suppose Example 1 and that we would like to obtain the willingness of Alice and Bob to accept changing their preferred actions for the conflicts found $C = \{Eve, Frank\}$. Suppose also that the tie strength between users are those given in Table 1. Table 2 shows all the willingness values for each of the conflicts and possible solutions. For instance, to calculate $\mathcal{W}(Alice, Eve)$, the mediator first calculates the item sensitivity and the relative importance of Eve as follows:

$$S_{Alice} = \frac{1}{\mid \mathcal{G}_{Alice} \mid} \sum_{G \in \mathcal{G}_{Alice}} \mathcal{T}_{Alice}(G) = \mathcal{T}_{Alice}(MyFriends) = 2$$

7. Note, however, that the calculations and meaning for sensitivity and relative importance are different and they may have different values for the same conflict, so this is why they are considered as two different dimensions for the willingness.

8. Given two n-dimensional vectors $\vec{p}$ and $\vec{q}$, the Canberra distance [31] is defined as:
$$d(\vec{p}, \vec{q}) = \sum_{i=1}^{n} \frac{\mid p_i - q_i \mid}{\mid p_i \mid + \mid q_i \mid}$$

and

$$I_{Alice}(Eve) =\mid \mathcal{T}_{Alice}(MyFriends) - \tau(Alice, Eve) \mid=\mid 2 - 1 \mid= 1$$

Then, the mediator estimates the willingness as follows:

$$\begin{aligned}
\mathcal{W}(Alice, Eve) &= \frac{1}{2} \cdot \left(\frac{\mid \delta - I_{Alice}(Eve) \mid}{\delta + I_{Alice}(Eve)} + \frac{\mid \delta - S_{Alice} \mid}{\delta + S_{Alice}} \right) \\
&= \frac{1}{2} \cdot \left(\frac{\mid 5 - 1 \mid}{5 + 1} + \frac{\mid 5 - 2 \mid}{5 + 2} \right) \\
&= \frac{1}{2} \cdot \left(\frac{4}{6} + \frac{3}{7} \right) \approx 0.55
\end{aligned}$$

We can see in Table 2 that the mediator would estimate Alice's willingness to grant Eve access to the item higher than Alice's willingness to grant Frank access to the item — recall Alice's preferred action for both Eve and Frank is to deny access, so the mediator estimates *willingness to grant access*. The reason for the estimated willingness is that, though the item seems not very sensitive for Alice ($S_a = 2$), Eve is closer to Alice than Frank, who seems not to be friend of Alice at all or be a very distant acquaintance because of a 0 tie strength. We can also see in Table 2 that the mediator would estimate Bob's willingness not to share with Eve to be lower than Bob's willingness not to share with Frank — recall Bob's preferred action for both Eve and Frank is to grant access, so the mediator estimates *willingness to deny access*. This is because Eve seems to have higher relative importance than Frank for Bob; i.e., Eve seems to be best friends with Bob (high tie strength), so it is plausible to believe Bob would definitely want to share with his best friend and would be unwilling to accept not sharing with her.

	Charlie	Dan	Eve	Frank
Alice	4	2	1	0
Bob	3	2	5	2

TABLE 1
Tie strength for Example 2, with $\delta = 5$ according to [24].

	Eve	Frank
Alice	0.55	0.43
Bob	0.34	0.71

TABLE 2
Willingness for Example 2.

5.2 Modelling Concessions

As suggested by existing research [3], [4], [5], negotiations about privacy in social media are collaborative most of the time. That is, users would consider others' preferences when deciding to whom they share, so users
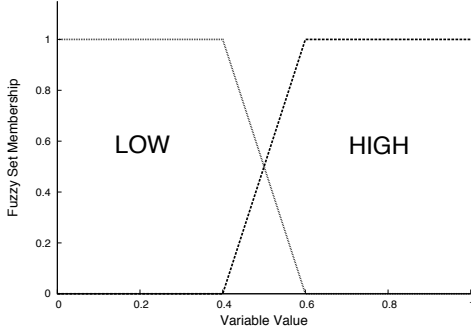
Fig. 1. Fuzzy sets low and high.

may be willing to concede and change their initial most-preferred option. Being able to model the situations in which these concessions happen is of crucial importance to propose the best solution to the conflicts found — one that would be acceptable by all the users involved. To this aim, the mediator models users' decision-making processes during negotiations based on the willingness to change an action (defined above) as well as on findings about *manual* negotiations in this domain, like the ones described in [3], [4], [5]. Users' decision making on continuous variables, like the willingness to accept an action, is commonly modelled using fuzzy sets that characterize intervals of the continuous variables [32]. Figure 1 depicts the intervals the mediator considers for the willingness to accept an action. This means that the mediator can consider the willingness to accept an action for a conflicting target user to be low or high. Based on this, the mediator considers fuzzy IF-THEN rules to model *concessions* in different situations as described below.

I do not mind (IDM) rule

Users are generally willing to accommodate others' sharing preferences [3], [4], so if they do not mind much about which action is finally applied, they will concede and accept applying the action that is not the most preferred for them. In particular, if the willingness to accept the action that is not the preferred one is high, then this may mean that the user would not mind much conceding and accepting that action for the conflicting target user. Assuming a negotiating user $a \in N$, and a conflicting target user $c \in C$, this concession can be formalised as the following fuzzy IF-THEN rule:

$$\textbf{IF } \mathcal{W}(a,c) \textbf{ IS } \texttt{high THEN } \texttt{concede} \qquad \text{(IDM)}$$

Note that concede means that user a would accept changing her initial most preferred action to reach an agreement. Thus, users that would initially prefer granting the particular conflicting target user access to the corresponding item would accept denying access, and users that would initially prefer denying the particular conflicting target user access to the corresponding item would accept granting access. For instance, Alice and

Bob could be depicted in a photo with very low sensitivity — e.g., a photo in which both Alice and Bob are depicted with a view to a monument — and both of them could have defined privacy policies for the photo so that all their friends can see it. Suppose that Charlie is friend of Alice but is distant acquaintance of Bob, so according to Alice's privacy policy Charlie should be granted access to the photo but according to Bob's privacy policy Charlie should not be granted access to the photo. However, given that the photo is not sensitive for Bob, Bob would probably accept sharing also with Charlie and solve the conflict.

I understand (IU) rule

Even when the willingness to change an action is low for some of the negotiating users, users do not want to cause any deliberate harm to their friends and will normally listen to their objections [4]. That is, if some of the negotiating users prefer denying a conflicting target user access and the willingness to grant access is low, then other users whose most preferred action for the target user is granting access and the willingness to accept denying is also low would *concede* and accept denying access to the conflicting target user. Indeed, considering others self-presentation online has been reported as a way of reaffirming and reciprocating user's relationships [4], [27]. Assuming a negotiating user $a \in N$, and a conflicting target user $c \in C$, this concession can be formalised as the following fuzzy IF-THEN rule:

$$\begin{aligned} \textbf{IF } \quad &\mathcal{W}(a,c) \textbf{ IS } \texttt{low} \quad \wedge \ v_a[c] = 1 \ \wedge \\ &\exists b \in N, \ \mathcal{W}(b,c) \textbf{ IS } \texttt{low} \quad \wedge \ v_b[c] = 0 \\ &\qquad\qquad\qquad\qquad \textbf{THEN } \texttt{concede} \quad \text{(IU)} \end{aligned}$$

For instance, Alice, Bob, and Charlie are depicted together in a photo in which Bob is clearly inebriated. Initially, Alice and Charlie might very much like to share the photo with friends because Alice, Bob and Charlie could agree they had a very good time together that day in which the photo was taken. However, Alice and Charlie would probably understand the privacy implications this may entail to Bob. Thus, if Bob opposes sharing the photo, Alice and Charlie would probably accept not sharing the photo.

No concession (NC) rule

For the other cases in which neither IDM nor IU applies, then the mediator estimates that a negotiating user would not concede and would prefer to stick to her preferred action for the conflicting target user. For completeness, this can be formalised as the following fuzzy IF-THEN rule assuming a negotiating user $a \in N$, and a conflicting target user $c \in C$:

IF $\mathcal{W}(a,c)$ **IS** low $\wedge$
$(v_a[c] = 0 \ \vee \ (\nexists b \in N : \mathcal{W}(b,c) \ \mathtt{IS} \ \mathtt{low} \ \wedge \ v_b[c] = 0))$
$\qquad\qquad\qquad$ **THEN** do not concede $\quad$ (NC)

For instance, when the willingness to accept granting grating access to the item is low, users very much seek to avoid sharing the item [22], because it can cause them a privacy breach; i.e., a sensitive item ends up shared with someone they would not like —e.g., in the example above, Bob would most probably not accept sharing the photo in which he appears inebriated with Alice and Charlie's friends because he might feel embarrassed about the photo and would prefer that no one sees it.

5.3 Computing Conflict Resolution

The mediator computes the solution for each conflict found by applying the concession rules defined above. The solution will be encoded into an action vector $\vec{o}$, so that $o[t]$ contains the action for target user t. If t is not conflicting, the mediator assigns to this target user the action shared by all negotiation users. If t is conflicting, the mediator assigns to $o[t]$ its proposal to solve the conflict. To this aim, the mediator executes Algorithm 2. In particular, for each conflicting target user t:

- If for all negotiating users, their willingness to accept changing their preferred action for the conflicting target user is high, then, according to concession rule IDM, the mediator assumes that all users are willing to concede if need be, so that the final action to be applied for target user t can be both grating and denying. In order to select one of these two actions, the mediator runs a modified majority voting rule (Lines 3-6). In particular, this function selects the action that is most preferred by the majority of users. In case that there is a tie — i.e., the number of users who prefer granting and the number of users who prefer denying is the same, then the one that uploaded the item is given an extra vote. Note that this function is only used if all the users have a high willingness to accept the action that is not the most preferred for them. That is, it does not really make much of a difference for them which action is finally taken, and all of them are willing to concede (change their preferred action) to reach an agreement.
- If there are users whose willingness to accept changing their preferred action for the conflicting target user is low (Lines 8-14), then the mediator considers two cases: (i) if there are at least two users with low willingness and different preferred actions, then, according to concession rule IU, the action to be taken should be denying the conflicting target user access to the item in question; (ii) otherwise, rule IDM applies so that the users that have high willingness will concede and the user/users who has/have low willingness will determine the action that is finally chosen as the solution.

The complexity of Algorithm 2 is $\mathcal{O}(|C|\times|N|^2)$. That is, for each conflict, we need to know for each negotiating agent what is her willingness, which can be calculated in constant time as the sensitivity would only need to be calculated once for all conflicts, and the relative importance of each particular conflicting user can be obtained in constant time. Note that in the very worst case; i.e., the very unlikely case that all target agents turn out to be in conflict, then $C = T$ and the complexity of Algorithm 2 would be $\mathcal{O}(|T| \times |N|^2)$.

Algorithm 2 Conflict Resolution

Input: N, $P_{n_1}, \ldots, P_{n_{|N|}}$, C
Output: $\vec{o}$
1: **for all** $c \in C$ **do**
2:
3: $\quad$ **if** $\forall n \in N$, $\mathcal{W}(n,c)$ is HIGH **then**
4: $\qquad$ $o[c] \leftarrow$ modified_majority$(P_{n_1}, \ldots, P_{n_{|N|}}, c)$
5: $\qquad$ **continue**
6: $\quad$ **end if**
7:
8: $\quad$ **if** $\exists a \in N$, $\mathcal{W}(a,c)$ is LOW **then**
9: $\qquad$ **if** $\exists b \in N$, $\mathcal{W}(b,c)$ is LOW $\wedge v_a[c] \neq v_b[c]$ **then**
10: $\qquad\quad$ $o[c] \leftarrow 0$
11: $\qquad$ **else**
12: $\qquad\quad$ $o[c] \leftarrow v_a[c]$
13: $\qquad$ **end if**
14: $\quad$ **end if**
15: **end for**

	Eve	Frank
Alice	HIGH	LOW
Bob	LOW	HIGH

TABLE 3
Fuzzy Memberships of willingness for Example 3.

Example 3: Suppose again Example 1 and consider the willingness values calculated in Example 2. Table 3 shows the fuzzy set membership for negotiating users Alice and Bob in case they would accept changing their most preferred action for the conflicting target users $C = \{\text{Eve}, \text{Frank}\}$. We can see that for Alice and Eve IDM rule applies, so that the mediator assumes that Alice would concede (in this case, to accept granting Eve access to the item). As Bob has willingness LOW to change his preferred action for Eve, then the action suggested by this user would be taken to solve the conflict, and the computed solution would be to grant Eve access to the item. Regarding Frank, we have a similar situation. In this case, the willingness is HIGH for Bob, so that IDM rule applies and Bob would concede. As there is only one negotiating user (Alice) with willingness LOW, then the action suggested by this user is taken to solve the conflict. Therefore, the solution to the conflict would be to deny Frank access to the item. The resulting action

vector for the item would be $\vec{o} = \{1, 1, 1, 0\}$; i.e., Charlie, Dan and Eve would be granted access to the item while Frank would be denied access to the item.

6 USER STUDY

The aim of this section is to compare the performance of our proposed mechanism to other existing approaches in terms of what users would do themselves manually in a number of situations. To this aim, we conducted the user study described below.

6.1 Method

We sought to explore situations with different degrees of sensitivity, as users' behaviour to resolve conflicts may be different depending on how sensitive items are. However, this would have involved participants sharing with us sensitive items of them. Participants sharing sensitive information in user studies about privacy in Social Media was already identified as problematic in related literature [23], as participants would always seem reluctant to share sensitive information, which biases the study towards non-sensitive issues only. Indeed, this reluctance to share information that may be sensitive with researchers during user surveys is not only associated with studies about privacy and Social Media, but it has also been extensively proven to happen in many other survey situations, including other scientific disciplines such as psychology [33]. A possible alternative to avoid this problem could be one in which participants just self-report how they behave when they experience a multi-party privacy conflict without asking for any sensitive information of them. However, the results obtained in that case may not match participants' actual behaviour in practice, as previous research on privacy and Social Media showed that there is a dichotomy between users' stated privacy attitudes and their actual behaviour [34]. As a trade-off between these two alternatives, we chose to recreate *situations* in which participants would be *immersed*, following a similar approach to [35], maximising actual behaviour elicitation while avoiding biasing the study to non-sensitive situations only. To this aim, we described a situation to the participants and asked them to immerse themselves in the situation by thinking they were a particular person in a particular photo that was to be shared through a Social Media site and that they were tagged in it, and participants showed very different individual privacy policies and concession decisions depending on the situation as detailed below. Each participant was presented with 10 different scenarios. Scenarios were different across participants as they were composed of: (i) one photo involving multiple users; and (ii) a conflict created based on the individual privacy policy the participant specified for the photo. As we had 50 participants (as detailed below), we were able to gather participant-specified data relative to 500 different scenarios. Photos referred to different situations (e.g., travelling, playing with friends, partying, dating, etc.)

and were of different sensitivities *a priori* — though the participants were asked to specify their privacy policy for the photo as their first task for each scenario (as detailed below), which was different according to how sensitive each photo was for each participant.

We developed a web application that presented the participants with the photos, stored the individual privacy policy they selected for each photo, generated conflicts, and stored whether or not participants would *concede* during a negotiation in the scenarios presented. For each scenario, participants completed the following two tasks using the application:

1) **Definition of the Individual Privacy Policy**. Each participant was asked to define her/his most preferred privacy policy for each photo.
2) **Conflict and Concession Question**. Once the participants defined their individual privacy policy for the photo, a conflict was generated. That is, we told the participants that one or more of the other people in the photo had a different most preferred action for one particular person, specifying the relationship type and strength the participant would have to this person. For instance, if the participant only wanted to share the photo with close friends, we told her/him that the other people in the photo wanted to share the photo with someone that was her/his acquaintance. Where multiple options were available to generate a conflict, we chose one of them randomly. Then, we asked participants whether or not they would concede and change their most preferred action for that person to solve the conflict with the other people depicted in the photo.

6.2 Participants

We recruited 50 participants via e-mail including university students, academic and non-academic staff, as well as other people not related to academia who volunteered to participate in the study. Participants completed the study online using the web application developed to that end (as detailed above). Before starting, the application showed the information to be gathered and participants needed to consent to continue. Table 4 summarises participants' demographics (gender, age, job), Social Media use (number of accounts in different Social Media sites, and frequency of use), and if they were concerned about their privacy in Social Media (Priv. concerned).

6.3 Results

The results gathered through the web application were compared to the results that would have been obtained if our proposed mechanism was applied to the scenarios and if state-of-the-art automated voting mechanisms were applied. To this aim, we looked at the privacy policy defined by the participant and the conflict generated by the application for each situation. This determined

Variable	Distribution
Gender	female (42%), male (58%)
Age	18-24 (18%), 25-30 (36%), 31-40 (24%), 41-50 (10%), 51-60 (6%), 60+(6%)
Job	Agriculture(4%), Arts(2%), Computers(26%), Design(6%),Education(16%),Engineering (10%), Management(4%),Media(2%), Research(14%),Sales(2%), Other(14%)
# accounts	0(4%),1(30%),2(18%),3(8%),4(12%),4+(28%)
Freq. of use	monthly- (18%), monthly (10%), weekly(10%), daily (26%), daily+(36%)
Priv. concerned	not much(36%), yes(26%), very much(36%)

TABLE 4
Participants' demographics, Social Media use, and privacy concern.

Fig. 2. Percentage of times each approach matched concession behaviour.

participants' most preferred action for the conflict (to be considered by our proposed mechanism and state-of-the-art voting mechanisms), as well as the willingness to change it (used to determine the concession rule our mechanism would apply in each case). In particular, we compared the results that would have been obtained applying our proposed mechanism to those that would have been obtained applying the general voting mechanisms used in state-of-the-art automated approaches:

- **Uploader overwrites** (UO), the conflict is solved selecting the action preferred by the user that uploads the item. This is the strategy currently followed by most Social Media Sites (Facebook, etc.).
- **Majority voting** (MV) [12], the conflict is solved selecting the action most preferred by the majority of the negotiating users.
- **Veto voting** (VV) [2], if there is one negotiating user whose most preferred action is denying access, the conflict is solved by denying access to the item.

Figure 2 shows the results for each of the above voting mechanisms as well as the results for our proposed mechanism for automated conflict resolution (labelled AR in the figure). In particular, it shows the percentage of times each mechanism matched participants' concession behaviour in the scenarios above. We can observe that our proposed mechanism AR clearly outperformed UO, MV, and VV. This is because these mechanisms lack enough flexibility to model actual user behaviour across different situations in this domain, as they only consider the most preferred action for each negotiating user as a vote without considering the particular situation. We can also observe that UO is very far from what users did themselves, which is mainly due to UO not being collaborative at all —i.e., the preferences of the other parties are not considered. MV performs a bit better than UO, but it is still far from what participants did themselves. This is mostly due to the situations in which even if the majority of users would like to share an item in the first instance, they could reconsider this if there is/are one/multiple user/(s) that would prefer not sharing because this could have privacy consequences for them.

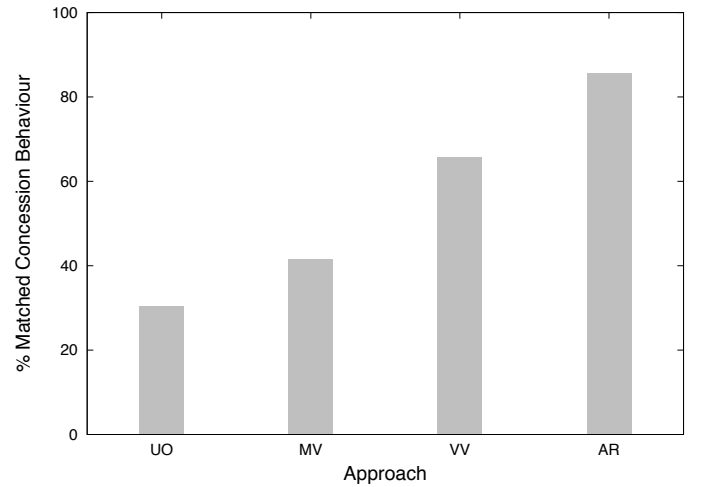We can also see in Figure 2 that VV performs better than UO and MV. This result confirms that negotiating

users are many times open to accept not sharing an item if this can cause privacy breaches to one of them — as also modelled in our proposed mechanism AR. However, VV is too restrictive to be suitable for all situations. This is because there are also situations in which the user/s whose most preferred action is denying access may not mind granting access due to many reasons. In these cases, VV would suggest solutions that mean losing sharing opportunities. For instance, as stated earlier, Alice and Bob could be depicted in a photo with very low sensitivity — e.g., a photo in which both Alice and Bob are depicted with a view to a monument — and both of them could have defined privacy policies for the photo so that all their friends can see it. Suppose that Charlie is friend of Alice but is distant acquaintance of Bob, so according to Alice's privacy policy Charlie should be granted access to the photo but according to Bob's privacy policy Charlie should not be granted access to the item. However, given that the photo is not sensitive for Bob, Bob would probably accept sharing also with Charlie. VV would not consider this concession, and the solution to solve the conflict would be not sharing with Charlie, so it would be a lost sharing opportunity and Alice may not even accept the solution. In contrast, our mechanism is able to adapt to the particular situation, being as restrictive as VV if needed but also considering the cases in which *concessions* about granting access are to happen —as the example above, in which the I do not mind (IDM) rule would have picked that Bob would concede, so that the final solution would be to share with Charlie (recall the item was not sensitive to Bob).

We also sought to explore more closely how each of the concession rules in our proposed mechanism contributed to its performance as well as how state-of-the-art voting mechanisms would work in each case. Table 5 shows for each concession rule the number of times that each rule would have been applied (# Instantiations) in the 500 situations and Figure 3 shows the performance of each approach broken down by the
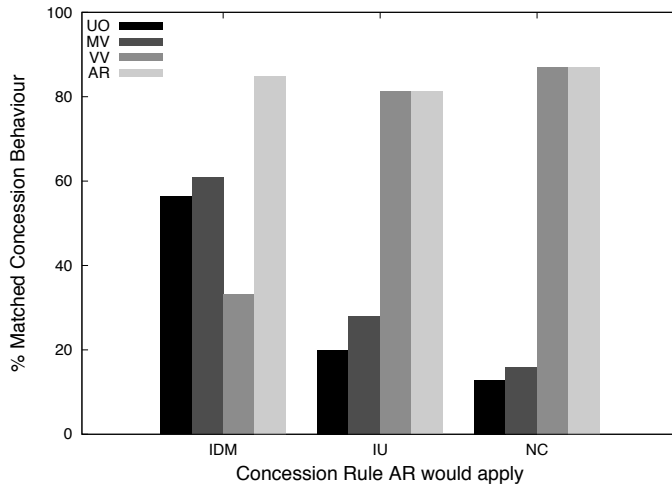
Fig. 3. Percentage of times each approach matched concession behaviour broken down by the concession rule AR would apply (IDM - I do not mind, IU - I understand, NC - No concession).

concession rule that would have been applied for each situation. We can observe that performance was similar across concession rules for our proposed mechanism AR; i.e., once a particular concession rule instantiated for a situation, it usually matched users' behaviour with respect to concessions. In particular, we observe that the three concession rules in our mechanism obtain better results than the state-of-art approaches. We can also observe that the performance of state-of-the-art voting mechanisms significantly varied according to the concession rule AR would apply. This confirms the fact that static ways of aggregating preferences (as those used in state-of-the-art voting mechanisms) are not desirable in this domain, because the concessions that may happen to resolve multi-party privacy conflicts clearly depend on the particular situation —as captured by the variables considered by AR's concession rules; i.e., individual privacy preferences of each user, the sensitivity of the item to be shared, and the relative importance of the conflicting target user.

Concession Rule	# Instantiations
I do not mind (IDM)	172
I understand (IU)	111
No concession (NC)	217
Total	500

TABLE 5
Number of times each AR concession rule would have been applied.

Finally, we sought to find any correlation that could exist between participants' data —demographics, social media use and privacy concern— and whether or not participants behaved according to the concession rule instantiated for each situation. To this aim, we calculated the information gain (IG) — i.e., the reduction in entropy — that each variable produced on whether the participant followed the corresponding rule or not once it was

	Rule 1		Rule 2		Rule 3	
	IG	CC	IG	CC	IG	CC
Age	0	0.04	0	0	0	-0.10
Gender	0.06	0.14	0.06	0.13	0	-0.06
Job	0.08	-0.18	0.08	-0.17	0.04	0.11
Studies	0	0.18	0	0.16	0	0.017
Freq. of use	0.03	0.05	0.06	0.08	0.02	-0.07
# Accounts	0	0.14	0	0.16	0	-0.05
Priv. Concern	0	0.13	0	0.16	0	0.04

TABLE 6
IGs and CCs for each rule based on participants' demographics, social media use, and privacy concern. All the correlation coefficients obtained were *not* statistically significant (i.e., $p > 0.05$).

instantiated, and the Pearson's correlation coefficient (CC). Table 6 summarises the values for each rule. IGs and CCs were negligible and not statistically significant. Thus, users' characteristics like the privacy concern, age, gender, profession, studies, and social media use did not have any significant effect on whether participants followed a concession rule once the rule was instantiated for a particular situation. Note, however, the particular concession rule instantiated in each situation depended on the individual privacy policy of each user, the sensitivity of the item for the user, and the relative importance of the conflicting target user as stated above, which may vary from participant to participant. The important thing is that once a rule was instantiated, the variables above did not influence whether the particular instantiated rule was successful in matching user behaviour or not. In other words, users' characteristics (e.g., demographics, privacy concern, etc.) may determine the individual privacy policies users choose, which in turn determine the rules that are instantiated for a given situation; but users' characteristics do not determine whether users' concession behaviour matches that of the rule instantiated. This suggests the mechanism proposed in this paper captures general user behaviour and would be able to adapt to both different situations and users.

7 RELATED WORK

Until now, very few contributions have considered the problem of resolving conflicts in multi-party privacy management for Social Media. Wishart et al. [10] proposed a method to define privacy policies collaboratively. In their approach all of the parties involved can define strong and weak privacy preferences. However, this approach does not involve any automated method to solve conflicts, only some suggestions that the users might want to consider when they try to solve the conflicts manually.

The work described in [11] is based on an incentive mechanism where users are rewarded with a quantity of numeraire each time they share information or acknowledge the presence of other users (called co-owners) who are affected by the same item. When there are

conflicts among co-owners' policies, users can spend their numeraire bidding for the policy that is best for them. Then, the use of the Clark Tax mechanism is suggested to obtain the highest bid. As stated in [13], users may have difficulties to comprehend the mechanism and specify appropriate bid values in auctions. Furthermore, users that earned much numeraire in the past — e.g., because they shared much information — will have more numeraire to spend it at will, potentially leading to unilateral decisions.

In [13] users must manually define for each item: the privacy settings for the item, their trust to the other users, the sensitivity of the item, and how much privacy risk they would like to take. These parameters are used to calculate what the authors call privacy risk and sharing loss on segments — they define segments as the set of conflicting target users among a set of negotiating users. Then, based on these measures all of the conflicting target users in each segment are assigned the same action. That is, all of the conflicts that a set of negotiating users have would be solved either by granting or denying access. Clearly, not considering that each individual conflict can have a different solution leads to outcomes that are far from what the users would be willing to accept. Moreover, due to how the privacy risk and sharing loss metrics are defined, solutions are likely to be the actions preferred by the majority of negotiating users, which can be many times far from the actual behaviour of users as shown in Section 6.

There are also related approaches based on voting in the literature [2], [12]. In these cases, a third party collects the decision to be taken (granting/denying) for a particular friend from each party. Then, the authors propose to aggregate a final decision based on one of the voting rules already been described in Section 6 — i.e., uploader overwrites (UO), majority voting (MV), and veto voting (VV). These approaches are static, in the sense that they always aggregate individual votes in the same way by following the same voting rule. Thus, these approaches are unable to adapt to different situations that can motivate different concessions by the negotiating users, which makes these approaches unable to match the actual behaviour of users many times, as shown in Section 6. Only in [14], the authors consider that a different voting rule could be applied depending on the situation. However, it is the user who uploads/posts the item the one who chooses *manually* which one of the voting rules (UO,MV,VV) to apply for each item. The main problem with this — apart from having to specify the voting rule manually for every item — is that the choice of the voting rule to be applied is unilateral. That is, the user that uploads the item decides the rule to apply without considering the rest of the negotiating users' preferences, which becomes a unilateral decision on a multi-party setting. Moreover, it might actually be quite difficult for the user that uploads the item to anticipate which voting rule would produce the best result without knowing the preferences of the other users.

Finally, the problem of negotiating a solution to multi-party conflicts, has also been very recently analysed from a game-theoretic point of view [36], [37]. These proposals provide an elegant analytic framework proposing negotiation protocols to study the problem and the kind of solutions that can be obtained based on well-known game-theoretic solution concepts such as the Nash equilibrium. However, as shown in [36], these proposals may not always work well in practice, since they are not able to capture the social idiosyncrasies considered by users in the real life when they face multi-party privacy conflicts, and users' behaviour is far from perfectly rational as assumed in these game-theoretic approaches — e.g., refer to [3], [4].

8 CONCLUSIONS

In this paper, we present the first mechanism for detecting and resolving privacy conflicts in Social Media that is based on current empirical evidence about privacy negotiations and disclosure driving factors in Social Media and is able to adapt the conflict resolution strategy based on the particular situation. In a nutshell, the mediator firstly inspects the individual privacy policies of all users involved looking for possible conflicts. If conflicts are found, the mediator proposes a solution for each conflict according to a set of concession rules that model how users would actually negotiate in this domain.

We conducted a user study comparing our mechanism to what users would do themselves in a number of situations. The results obtained suggest that our mechanism was able to match participants' concession behaviour significantly more often than other existing approaches. This has the potential to reduce the amount of *manual* user interventions to achieve a satisfactory solution for all parties involved in multi-party privacy conflicts. Moreover, the study also showed the benefits that an adaptive mechanism like the one we presented in this paper can provide with respect to more static ways of aggregating users' individual privacy preferences, which are unable to adapt to different situations and were far from what the users did themselves.

The research presented in this paper is a stepping stone towards more automated resolution of conflicts in multi-party privacy management for Social Media. As future work, we plan to continue researching on what makes users concede or not when solving conflicts in this domain. In particular, we are also interested in exploring if there are other factors that could also play a role in this, like for instance if concessions may be influenced by previous negotiations with the same negotiating users.

Finally, in this paper we focused on *centralised* Social Media, such as Facebook, Twitter, Google+ and the like, because this type of Social Media is the most widespread. However, another possible future line of research could be that of investigating *distributed* negotiation protocols without a centralised mediator. This could be of particular interest for the emerging *decentralised* Social Media

infrastructures such as Friendica [38] and Diaspora [39], in which the information about the privacy policies and relationships among users would be distributed among the different instances that make up the social network. Even for such distributed negotiation protocols, most of the concepts and formalisms we have presented in this paper would still be useful. For instance, the willingness measures and the concession rules could be used locally by each social media instance during the execution of the distributed negotiation protocol to decide whether the negotiating users hosted in that particular instance of the social network would be willing to concede or not.

REFERENCES

[1] Internet.org, "A focus on efficiency," http://internet.org/efficiencypaper, Retr. 09/2014.
[2] K. Thomas, C. Grier, and D. M. Nicol, "unfriendly: Multi-party privacy risks in social networks," in *Privacy Enhancing Technologies*. Springer, 2010, pp. 236–252.
[3] A. Lampinen, V. Lehtinen, A. Lehmuskallio, and S. Tamminen, "We're in it together: interpersonal management of disclosure in social network services," in *Proc. CHI*. ACM, 2011, pp. 3217–3226.
[4] P. Wisniewski, H. Lipford, and D. Wilson, "Fighting for my space: Coping mechanisms for sns boundary regulation," in *Proc. CHI*. ACM, 2012, pp. 609–618.
[5] A. Besmer and H. Richter Lipford, "Moving beyond untagging: photo privacy in a tagged world," in *ACM CHI*, 2010, pp. 1563–1572.
[6] Facebook NewsRoom, "One billion- key metrics," http://newsroom.fb.com/download-media/4227, Retr. 26/06/2013.
[7] J. M. Such, A. Espinosa, and A. García-Fornes, "A survey of privacy in multi-agent systems," *The Knowledge Engineering Review*, vol. 29, no. 03, pp. 314–344, 2014.
[8] R. L. Fogues, J. M. Such, A. Espinosa, and A. Garcia-Fornes, "Open challenges in relationship-based privacy mechanisms for social network services," *International Journal of Human-Computer Interaction*, no. In press., 2015.
[9] S. Vihavainen, A. Lampinen, A. Oulasvirta, S. Silfverberg, and A. Lehmuskallio, "the clash between privacy and automation in social media," *Pervasive Computing, IEEE*, vol. 13, no. 1, pp. 56–63, 2014.
[10] R. Wishart, D. Corapi, S. Marinovic, and M. Sloman, "Collaborative privacy policy authoring in a social networking context," in *POLICY*. IEEE, 2010, pp. 1–8.
[11] A. Squicciarini, M. Shehab, and F. Paci, "Collective privacy management in social networks," in *WWW*. ACM, 2009, pp. 521–530.
[12] B. Carminati and E. Ferrari, "Collaborative access control in online social networks," in *IEEE CollaborateCom*, 2011, pp. 231–240.
[13] H. Hu, G.-J. Ahn, and J. Jorgensen, "Detecting and resolving privacy conflicts for collaborative data sharing in online social networks," in *Proc. ACSAC*. ACM, 2011, pp. 103–112. [Online]. Available: http://doi.acm.org/10.1145/2076732.2076747
[14] H. Hu, G. Ahn, and J. Jorgensen, "Multiparty access control for online social networks: model and mechanisms," *IEEE TKDE*, 2013.
[15] B. Carminati, E. Ferrari, and A. Perego, "Enforcing access control in web-based social networks," *ACM TISSEC*, vol. 13, no. 1, p. 6, 2009.
[16] P. Fong, "Relationship-based access control: protection model and policy language," in *Procs. ACM CODASPY*. ACM, 2011, pp. 191–202.
[17] J. M. Such, A. Espinosa, A. Garcia-Fornes, and C. Sierra, "Self-disclosure decision making based on intimacy and privacy," *Information Sciences*, vol. 211, pp. 93–111, 2012.
[18] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *WWW*. ACM, 2010, pp. 351–360.
[19] A. C. Squicciarini, S. Sundareswaran, D. Lin, and J. Wede, "A3p: adaptive policy prediction for shared images over popular content sharing sites," in *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*. ACM, 2011, pp. 261–270.
[20] G. Danezis, "Inferring privacy policies for social networking services," in *Proceedings of the 2nd ACM workshop on Security and artificial intelligence*. ACM, 2009, pp. 5–10.
[21] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1701–1708.
[22] M. Sleeper, R. Balebako, S. Das, A. L. McConahy, J. Wiese, and L. F. Cranor, "The post that wasn't: exploring self-censorship on facebook," in *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 793–802.
[23] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor, "I regretted the minute i pressed share: A qualitative study of regrets on facebook," in *Proceedings of the Seventh Symposium on Usable Privacy and Security*. ACM, 2011, p. 10.
[24] R. L. Fogués, J. M. Such, A. Espinosa, and A. Garcia-Fornes, "BFF: A tool for eliciting tie strength and user communities in social networking services," *Information Systems Frontiers*, vol. 16, no. 2, pp. 225–237, 2014.
[25] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *CHI*. ACM, 2009, pp. 211–220.
[26] E. Gilbert, "Predicting tie strength in a new medium," in *CSCW*. New York, NY, USA: ACM, 2012, pp. 1047–1056. [Online]. Available: http://doi.acm.org/10.1145/2145204.2145360
[27] K. Green, V. J. Derlega, and A. Mathews, "Self-disclosure in personal relationships," in *The Cambridge Handbook of Personal Relationships*. Cambridge University Press, 2006, pp. 409–427.
[28] D. J. Houghton and A. N. Joinson, "Privacy, social network sites, and social relations," *JTHS*, vol. 28, no. 1-2, pp. 74–94, 2010.
[29] J. Wiese, P. Kelley, L. Cranor, L. Dabbish, J. Hong, and J. Zimmerman, "Are you close with me? are you nearby? investigating social groups, closeness, and willingness to share," in *UbiComp*. ACM, 2011, pp. 197–206.
[30] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi, "Talking in circles: selective sharing in google+," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2012, pp. 1065–1074.
[31] G. N. Lance and W. T. Williams, "Mixed-data classificatory programs i - agglomerative systems," *Australian Computer Journal*, vol. 1, no. 1, pp. 15–20, 1967.
[32] L. A. Zadeh, "The calculus of fuzzy if/then rules," in *Proceedings of the Theorie und Praxis, Fuzzy Logik*. Springer-Verlag, 1992, pp. 84–94.
[33] R. Tourangeau and T. Yan, "Sensitive questions in surveys." *Psychological bulletin*, vol. 133, no. 5, p. 859, 2007.
[34] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in *Privacy enhancing technologies*. Springer, 2006, pp. 36–58.
[35] C. Mancini, Y. Rogers, A. K. Bandara, T. Coe, L. Jedrzejczyk, A. N. Joinson, B. A. Price, K. Thomas, and B. Nuseibeh, "Contravision: exploring users' reactions to futuristic technology," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 153–162.
[36] H. Hu, G.-J. Ahn, Z. Zhao, and D. Yang, "Game theoretic analysis of multiparty access control in online social networks," in *Proceedings of SACMAT '14*. New York, NY, USA: ACM, 2014, pp. 93–102. [Online]. Available: http://doi.acm.org/10.1145/2613087.2613097
[37] J. M. Such and M. Rovatsos, "Privacy policy negotiation in social media," *arXiv preprint arXiv:1412.5278*, 2014.
[38] *Friendica*, (accessed April 28, 2014), http://friendica.com/.
[39] *Diaspora*, (accessed February 3, 2014), https://diasporafoundation.org.
[40] J. M. Such and N. Criado, "Adaptive conflict resolution mechanism for multi-party privacy management in social media," in *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. ACM, 2014, pp. 69–72.

APPENDIX A
APPLYING THE MECHANISM TO REBAC

To showcase that our proposed mechanism can also be used in conjunction with other access control models, we include in this appendix the formalisation needed

to use relationship-based access control (ReBAC) [15], [16], [17] in conjunction with our mechanism. ReBAC defines individual privacy preferences differently from the group-based privacy policies considered in Def. 1, so it needs adapted definitions of the action function (Def. 2), item sensitivity (Def. 6), and relative importance (Def. 7), which are the functions that depend on how individual privacy preferences are expressed. Note that a preliminary formalisation of a ReBAC approach to solve multi-party conflicts appeared in [40], but the following is new in accordance to the conflict resolution mechanism presented in this paper.

ReBAC privacy policies usually consider different relationship types $R = \{r_1, \ldots, r_l\}$ — e.g., family, friends, colleagues, etc.; and define a mapping $r : U \times U \to R$ so that $r(a, b)$ is the relationship type between users a and b. Privacy policies in most ReBAC approaches [15] also consider the strength of the relationships that a user has to other users. A ReBAC privacy policy can be defined as follows:

Definition 8: A relationship-based privacy policy P is a tuple so that $P = \langle \theta_1, \ldots, \theta_{|R|}, E \rangle$, where $\theta_j \in \{0, \ldots, \delta\}$ is the relationship strength threshold for the relationship type $r_j \in R$, and $E \subseteq U$ is the set of exceptions to the policy.

The action function (Def. 2) would be then defined as:

Definition 9: Given a user $n \in N$, her relationship-based privacy policy $P_n = \langle \theta_1, \ldots, \theta_{|R|}, E \rangle$, and a user $t \in T$; we define the action function as:

$$act(P_n, t) = \begin{cases} 1 & \text{if } \tau(n, t) \geq \theta_{r(n,t)} \wedge t \notin E \\ 1 & \text{if } \tau(n, t) < \theta_{r(n,t)} \wedge t \in E \\ 0 & \text{otherwise} \end{cases}$$

Item sensitivity (Def. 6) for the case of relationship-based privacy policies would be defined as follows:

Definition 10: Given a user $n \in N$, relationship types R, and n's ReBAC privacy policy P_n for an item; the sensitivity of the item for n is defined as follows:

$$\mathcal{S}_n = \frac{1}{|R|} \sum_{r \in R} P_n.\theta_r$$

Finally, the last definition that would need to be adapted depending on the access control model is the relative importance of the conflicting target user (Def. 7), which can be defined for ReBAC as follows:

Definition 11: Given a user $n \in N$, relationship types R, n's ReBAC privacy policy P_n for an item, and a conflicting target user $c \in C$; the relative importance of c for n, denoted as $I_n(c)$, is defined as:

$$I_n(c) = |P_n.\theta_{r(n,c)} - \tau(n, c)|$$