Using Stratified Privacy for Personal Reputation Defense in Online Social Networks[*]

Hasan M. Jamil
Department of Computer Science
University of Idaho, USA
jamil@uidaho.edu

ABSTRACT

Personal reputation in online social networks is fundamentally different from privacy and has significantly different traits than enterprise reputation. We argue that personal reputation cannot be effectively managed and defended using contemporary online social networks privacy tools. In this paper, we propose a novel model for personal reputation management based on the concept of *stratified privacy*. We show that users of stratified privacy are better prepared to combat online harassment and reputation damage than traditional online social networks privacy based models, and are able to limit the damage to a minimum using the proposed tools. We also propose a declarative language, called *Green*Ship, that can be easily codified into a Facebook like convenient interface for social interaction for reputation defense in intuitive and flexible ways.

CCS Concepts

•**Information systems** → **Database design and models; Social networking sites;** *Query languages for non-relational engines; Reputation systems;* •**Security and privacy** → Human and societal aspects of security and privacy;

Keywords

Online social networks, stratified privacy, personal reputation, reputation defense, social contracts, content copyright

1. INTRODUCTION

The growth of social networks has had significant impact on how we interact, relate and at times, often avoid our families, friends and peers. Sites such as Facebook, Twitter, LinkedIn, Flickr, LivingSocial, Groupon and Pinterest have fundamentally changed the way we stay connected, improved our social presence and how we do commerce. In this virtual existence, our reputation and the way we maintain or improve it must necessarily change too. The Oxford English dictionary defines reputation as "the beliefs or opinions that are generally held about someone or something." For a person, or a user, an online reputation is her publicly held social evaluation and impressions based on her behavior, what she posts, and what others (such as individuals, groups, and recommendation services) share about her on the internet. The greater one's reputation, the more social influence and circle of friends she will have. Reputation thus is the currency for attracting larger number of online friends and followers, gaining access to desired circles of web communities, and influencing a significant section of online users [14, 26]. While there are a growing number of online reputation management systems and a large body of research on enterprize reputation management, such efforts are virtually non existent for personal reputation management.

There are a growing number of incidents of cyber-attacks on individuals aimed to destroy their online, and transitively, their real world reputation and life. As social networks become more and more pervasive in our lives, an increasing number of online users are falling victim to cyber-bullying, harassment, online grooming and reputation damage. Such attacks usually have significant mental health [21], financial, and professional implications [2, 6], and often lead to suicides and homicides [27]. The psychological and social impacts on individuals due to diminished social reputation is undeniably devastating, especially when it is damaged deliberately and maliciously by a so called "friend." Though bullying, harassment and reputation attacks tend to be more common to younger users [25, 35], no one is outside such potential dangers. While academic validations are still being compiled, some credible statistics suggest that 1) cyber-bullying and reputation damage impact about 60% online users [4, 11] and about half of all young population in USA, UK, Germany and France, and 2) harassers prefer social networking sites such as Facebook as their weapon [9, 10]. The scarcity of technology to combat personal reputation damage has prompted policy makers, program designers and mental health practitioners to educate users of potential pitfalls of online social networks, how to avoid being a victim, and how to defend themselves from attacks when they happen [3, 5] while legal enforcement and judicial apparatuses are slowly being put in place.

1.1 Defamation and Invasion of Privacy

A recent Pew research [28] offers a detailed exposé of online behavior of younger population and how their digital footprint is impacting their lives. This and other studies

[*]Research supported in part by the National Science Foundation grant DRL 1515550.

SAC 2017, April 03 - 07, 2017, Marrakech, Morocco

ISBN 978-1-4503-4486-9/17/04...$15.00
DOI: http://dx.doi.org/10.1145/3019612.3019813

show while the psychological precepts are varied, the ultimate anti-social act many online users commit to bully, harass, stalk, blackmail or torment their victims commit acts of defamation, or privacy violation. While there are laws in almost every country to hold defamatory acts liable, privacy violation laws are slowly emerging. In USA, only about 27 states have laws to protect its citizens against willful violation of one's privacy. The fundamental difference between the two is that defamatory statements (verbal or written) are false, while violation of privacy statements are true. On social networks, both are used as tools, and more often than not, these are privacy violation.

Under the privacy violation laws, "one may be sued for the dissemination of intimate information about a person, even if true, under the privacy tort of public disclosure of private facts, sometimes referred to as the embarrassing-facts tort or private-facts tort. The Restatement (Second) of Torts defines this cause of action as a publication of private information that (a) *would be highly offensive to a reasonable person* and (b) *is not of legitimate concern to the public*. In other words, the disclosure of very personal information, a disclosure unjustified by the newsworthiness, or lack thereof, of the information is an invasion of privacy. Note that the embarrassing information revealed must be private, meaning it is not in the public domain or otherwise generally known" [7]. The offence covers acts both on or offline, and includes the uploading of images or texts on the internet, sharing by text and e-mail, or showing someone a physical or electronic content of private nature.

Unfortunately, privacy violation is one of the most potent tools abusers use to damage reputation of their friends. This means that the content they use were voluntarily shared, and are true. Sensitive private communications, sexually explicit discussions, private images and videos, known generally as *revenge porns*, are all used to attack people. Revenge porn is the sharing of private, sexual material, either photos or videos, of another person without their consent and with the purpose of causing embarrassment or distress. The images are sometimes accompanied by personal information about the subject, including their full name, address, phone number and links to their social media profiles. Surveys such as in [10, 28] and academic research [13, 29] find that young people, senior citizens and technically challenged social net users share sensitive information the most, and among them, mostly young people, people with a social reputation and people after a romantic breakup are most vulnerable to such attacks. Unfortunately, such attacks cannot be foreseen or prevented using the current privacy tools supported by social networks such as Facebook.

1.2 The Telltale Signs of Abuse Susceptibility

The leading social networking system Facebook recently announced that they now have about 1.23 billion monthly active users of which 757 millions login to Facebook daily and 945 millions use Facebook on mobile devices. Among these, it is estimated that about 90 million Facebook accounts are fake. 63% of all Facebook profiles are publicly visible, and 55% of teenagers share information with the general public. Some estimates suggest that a) 30% of teenagers disclose something very embarrassing and harmful on Facebook to others including videos, photos, or simple rumors, b) 30% of teens in the USA are stalked by and have received friend requests from complete strangers, c) 83% regularly check the profiles of their previous intimate partners and 75% will stalk the new partners of their former partners [20], d) about 70% will use a friend's profile to stalk their former partners who blocked them, and finally e) about 16% of all social media stalking takes place on Facebook.

Putting these statistics together paints a serious threat that we are only beginning to understand and grapple with. Surveys and resources such as [2, 3, 10, 28] suggest that most people underestimate the threats. Most users believe that setting everything to private protects them and completely ignore the threats that the online "friends" pose. Damage takes place before they begin to realize what is happening, leaving them little or no time to backtrack. Our position in this paper is that possibly it is time to begin with the assumption that we do not trust anyone and thus we should have the ability to retract the contents we shared when needed, and limit the damage by partitioning and containing potential threats within a confined online space.

2. NEED FOR STRATIFIED PRIVACY

Imagine a social network user *Joe*, a young man, with five online friends – *Pierre, Max, Odelia, Pria* and *Dan* as shown in figure 1. His dad *Pierre* and his dad's friend *Max* are both family and thus his image to them matters to him the most. His real life friends are *Nina, Clint* and *Moira* with whom he shares a significant pool of secrets. His girlfriend *Sue* is traditional and believes in simple values although she is the jealous type. Being young and not thinking like *Sue* does, he flirts with his online friends *Odelia* and *Pria* often.

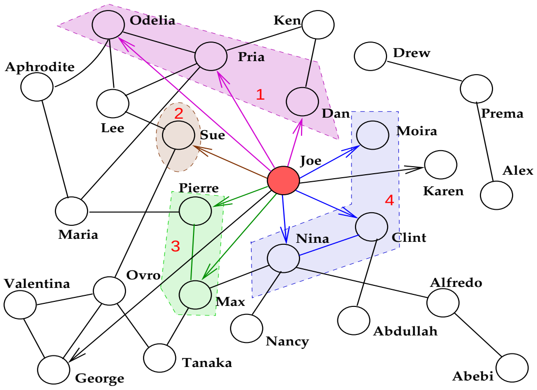

Figure 1: Joe's privacy sphere and his partitioned friendship network.

Joe wants to maintain his image and "reputation" as a smart gentleman, so he wants to partition his online behavior and disclosures in such a way that *Sue, Odelia* and *Pria* aren't aware of his indiscretions, and his relationship with *Sue* isn't threatened[1]. Above all, he does not want his dad and his dad's friend to know about any of his behavior. He

[1] This is not to condone *Joe*'s online behavior or to take a position on it morally. It is about what privacy should mean and how one can safeguard her online privacy and reputation leaving the moral decision with herself. It is possible to imagine equally compelling scenarios in other settings where users would like to partition their interactions. For example, a person in an abusive family who would like to interact with others in private, or a child keeping in touch with two warring parents or stepparents.

does not care so much if his real friends *Nina, Clint* and *Moira* know about his escapades. *Joe*'s fondness for others eventually spills over and is exposed through some of his posts, comments and other behaviors he engaged in online, and *Sue* begins to stalk him and his love interests. Question is, how can he engage in his online interactions safely, discretely and without any loss in reputation.

2.1 Personal Reputation Management

Joe can successfully hide his relationship and interaction with all three – *Sue, Odelia* and *Pria* – and make them mutually exclusive. To do so on Facebook, he can never post anything publicly, and keep all three in separate custom lists and post individually in these lists. Even if it is the exact same post, he will have to make the post three times separately and maintain all three posts to interact with them individually. Mixing the lists in any form will make him lose the exclusivity he seeks and his friends will know who else is in the custom list, and hence, his relationships with them. Additionally, maintaining exclusive interactions also will help him limit the damage any one of them may cause by spreading rumors to the people with whom he values his relationships the most, such as *Pierre* and *Max*. This is because *Sue, Odelia* and *Pria* will never know the existence of his relationships with the rest, and therefore cannot reach them. This is provided he never mixes the lists, and never posts anything publicly. We would like to remind the reader that in this model, we are summarily concerned about relationship disclosure, as opposed to content disclosure, by recognizing the fact that just hiding the friends list does not really protect relationship privacy.

2.2 Scalability and Inflexibility

The current Facebook approach available to *Joe*, as discussed above, is not truly scalable beyond a few friends. With an increasing number of friends, he is more likely to be vulnerable to exposing himself to potential abuse. This inflexible approach can be made more manageable if Facebook could recognize that even if the custom list consisted of other lists, the members in the lists are mutually disjoint, and "friends" actually meant all of *Joe*'s friends in all his lists. Then he could just post one message for all his friends, and yet maintain the exclusivity of the lists. As a further adaptation of "public" posts, *Joe* could post a message publicly to mean that friends in two lists would never see his interactions with the others. Thus, he could share his messages with the rest of the world, and hide his interactions with members in his mutually exclusive lists. We call this approach to hiding relationship disclosure with disjointed friends list and interaction exclusivity for common posts across lists *stratified privacy*.

2.3 Content Ownership and Retraction

Although privacy policies in Facebook have changed over the years, some policies appear to be stable. For example, a post and all the comments on the post are owned by the post-er, and the comment is owned by the commentator. While only the owners are allowed to edit or delete the content, the poster is allowed to delete any of the comments. Deleting a higher level comment, or the post, also deletes the comments and replies below. A shared content is owned by the poster, and is removed from other's timelines when the owner makes it unavailable in some manner – by deleting the content or by changing visibility. But, the "likes" are owned solely by the likers, and cannot be removed by the posters. In other words, by making it visible, the poster grants irreversible permissions to his "friends." Even blocking a friend does not change that permission. Blocking also does not remove the comments on a post, its just makes the comment and post invisible to the respective parties.

One of the most important contents in social networks in the context of reputation are chat logs. This is where users risk their reputation the most, yet they do not have the retraction authority over the content they author. In Facebook, a chat log is considered to be the property of the profile owner. In other words, if *Joe* chatted with *Sue*, the chat log at *Sue*'s end is owned by her, and the one at *Joe*'s end is owned by him. He cannot retract what he authored in a session, which is incidentally allowed on Skype. Deleting *Joe*'s log will not delete *Sue*'s. That also means, *Sue* (or *Joe*) can control the retention period of the chats, and use a sensitive chat log to cause damage to *Joe* (or *Sue*). The only way (and possibly an illegal way) *Joe* can get his chat log removed from *Sue*'s archive is by forcing *Sue* or other friends to report him as an abuser by committing acts that violate Facebook's use term and forcing Facebook to remove the chat log as offensive material. This cannot happen unless *Joe* somehow causes *Sue* to report him.

These complicated retraction options, and no real control over the content a user authors, leave open the door for real and tangible abuse, and combined with the relationship disclosures, make a perfect environment for reputation damage. Engaging themselves in some form of online sexual acts, *Joe* or *Sue* could potentially use their video clips as "revenge porn" and cause serious damage. Revenge porn, embarrassing and objectionable chat logs, sensitive private posts are tools that people use on Facebook to threaten, bully, stalk, abuse and torment their victims, and Facebook has no real defense mechanism to effectively help its users.

3. RELATED RESEARCH

While there is increasing interest in developing methods to help enterprizes [24, 36] and other social institutions to [32] improve their online image and reputation, the needed research is virtually non-existent. Social network systems take the position that reputation management is orthogonal to privacy, and users should be able to protect themselves by judiciously sharing sensitive information using the privacy tools they support. But applications and apps such as "RelationBook" [8] which informs a user which of their Facebook friends are currently single, "Diesel Cam" [31] in Spain that helps share trial/fitting room pictures with friends to receive immediate fashion feedback, and "Bang With Friends" [1] which aims to identify friends interested in sexual relationships undermine such simplistic wisdom.

To combat the risks, and to protect unsuspecting users, monitoring systems such as "Child Exploitation and Online Protection Centre" [16], "Net Nanny" [17], adult and pornographic content detection systems such as "NuDetective" [18], and even automatic cyber-bullying and other forms of abuse monitoring and reporting systems [33, 34] have been proposed. But these systems and approaches do not take into account that vast majority of online users (about 80%) actually do not want to be monitored and thus, find a way to circumvent them [30]. We believe, a better alternative is to offer effective privacy choices to the end users, and provide tools that will enable them to enter into *negotiated friendships* in which trust is earned, and backtracking is possible without harm. But when friendships do break down, malicious attacks can be contained and damage minimized.

4. PERSONAL REPUTATION V. PRIVACY

We believe that reputation damage of an individual is context sensitive. By that we mean that the perception of reputation damage depends on the type of sensitive information that is used to harass and defame someone and to whom, and there is no concrete characterization of either except that to cause most damage, one must use her most sensitive information to negatively impact her in the community with whom she values her relationships the most. Lesser the sensitivity or farther the distance of the relationships, less severe is the perceived damage. From this standpoint, we believe that limiting sensitive information to less trust-worthy friends and retracting shared information when needed is one effective way to protect content privacy. But when complete retraction is not possible, denying the harasser online access to friends who matter the most is another way to limit the damage based on stratified privacy.

We claim that the concept of stratified privacy is novel and introduces the definition of personal reputation for the first time. Until now, reputation basically has been considered synonymous to content privacy and thus, depends largely on whom sensitive information is shared with, and not how the shared information is used to harass, defame or bully the victim. We start by recognizing the fact that other than true perpetrators, most online relationships start with implicit trust and along the way someone betrays it and turns into an abuser. Once they do, they often find their victims ill prepared to combat their abuse. This is because research show that most Facebook users do not understand many powerful privacy tools it provides for its users to responsively share sensitive information [15, 37], with mostly using just the default settings. Although people care significantly about their online reputation [26], most are actually careless about safeguarding it [14, 28]. Thus a comprehensive model and a set of tools for reputation management has become an imperative orthogonally to net privacy awareness [3] efforts and tools to understand privacy setting implications [12, 19].

5. *GREEN*SHIP: A PERSONAL REPUTATION MANAGEMENT MODEL

As discussed earlier, most social network reputation attacks leverage information users share voluntarily with friends by spamming other friends with saved images, conversations or documents, real or engineered. The model of reputation management we introduce is aimed at denying the abuser the tools to spread information to users related to the victim, and limiting access to such damaging information.

We introduce our reputation model, called *Green*Ship, which stands for green or safe friendship, elucidated using the example in figure 1 and the definition of stratified privacy. In *Green*Ship, we recognize different types of friendships – global, standard, close and exclusive – all generally known as friends. However, these classifications carry privacy implications that are not possible in Facebook directly. In *Green*Ship, we view interactions among friends and users through posts and chats as a negotiated contract. This negotiation mutually agrees on the status of visibility, retention and retraction of the shared contents. Users have the choice to enter into such contracts, and interactions ensue only if an agreement is reached. Global search, newsfeed generation and timeline creation follow the privacy rules of *Green*Ship and only displays what is accessible in the context of a user.

While classification of friends into categories is optional, once classified, content sharing is impacted significantly, as is friend recommendations. First of all, everyone can receive anyone as her recommended non-mutual friend simply because they are considered a match based on some criteria other than their friendship with someone, e.g., interests, alumni status, location and so on. For example, *Prema* can be recommended as a possible non-mutual friend to *Joe, Pria, Aphrodite* or *Nina.* The only restriction on recommendation is that a friendship between two people cannot be revealed to a third friend if the disclosure is not intended by one of the first two. Users express this intention of non-disclosure by including their friendship information in the protected categories – standard, close and exclusive. Thus, two friends in two partitions will never be a mutual friend recommendation. For example, *Nina* or *Clint* will not be recommended to *Pierre* as a mutual friend of *Joe*, but *Max* will be recommended to *Clint* as a mutual friend of *Nina* even though he does not belong to the same partition. Note that *Joe*'s friend *Karen* is not in any group, and as such if *Maria* and *Joe* ever become friends, *Maria* can be a mutual friend with *Karen* but not *Pierre*, and no disclosure restriction among them will apply. Recall that the purpose of friendship partitioning or grouping is to separate, isolate and hide *Joe*'s relationships with them and to limit the damage a disgruntled friend can do by spreading damaging information to people to whom it matters. Thus the most damage a person can do is spread information within the group. *Green*Ship provides *Joe* with choices and tools he can use to judiciously create or modify group membership to limit the damage his friends can possibly do.

5.1 Close Friends

While everybody in *Green*Ship are friends, we extend the notion of "close friends and family" to characterize them as most trusted. In figure 1, *Pierre* and *Max* in group 3, shown in green, are two such friends of *Joe*, our main character shown in red. The friends of friend network of *Joe* for this group is thus the set {*Maria, Tanaka, Nina*}. Being the most trusted friends, they will be able to access all content (posts, images, comments and likes) *Joe* designates as global for friends or for this group only, including any specific subgroup of this group. The members of this group will be aware of each other and will be able to share comments and receive mutual friend suggestions even if *Joe* made the group invisible. For example, *Pierre* can receive *Max* as a mutual friend suggestion and become friends as shown by the solid edge between them. Similarly, *Maria* and *Tanaka* can be recommended as mutual friends to *Joe*. *Joe*'s group 2 friend *Sue* is also designated as a close friend, but being in a group by herself, she isn't aware of any of *Joe*'s friends in the other groups. Thus, she will not be able to access content *Joe* shares with any other group members, including his other close friends in group 3. But, she can receive *Pierre* and *Max* as mutual friend suggestions as a distinctive privilege of being in a close friend group.

5.2 Exclusive Friends

The least trusted friends of *Joe* are his group 1 friends *Odelia, Pria* and *Dan*, known as the exclusive friends. The main purpose of keeping them in an isolated category is to limit possible reputation damage they may cause in the future, should they so choose in the future. In addition to content retraction, they have no access to any of *Joe*'s friends. One of the differences with trusted group 3 friends is that *Odelia, Pria* and *Dan* are not aware of each other's friendship with *Joe*, even though they may be friends with each other themselves. For example, *Pria* and *Odelia* are friends as shown, but are not aware of each other's friendship with *Joe*. Essentially, these relationships are mutually exclusive as far as *Joe* is concerned, and thus they will not share anything among them what they share with *Joe*.

Friendship recommendations of exclusive friends are also different than standard and close friends of *Joe*. Since our goal is to keep mutual exclusivity, *Pria* (and *Odelia*) will never be recommended as a mutual friend of *Joe* to *Dan* (and vice versa), but *Pria* will be as a mutual friend of *Ken*. However, if *Joe*'s global friends are invisible, his group 1 friends will never receive *Karen* as their mutual friend suggestion, and *Karen* will never receive them regardless of her visibility status[2]. Note that Facebook supports a *restricted* friendship category allowing its members access to only public contents. Evidently, despite apparent similarity between the two, the semantics of exclusive friends in *Green*Ship is entirely different.

5.3 Standard and Global Friends

The Facebook equivalent of friends in *Green*Ship are the global friends, i.e., the unmanaged friends. In figure 1, *Karen* and *George* are two such friends of *Joe*, who are not part of any protected groups of friends and form a group by themselves. Similarly, standard friends are like global friends but in a protected list. Standard friends are thus similar to Facebook custom friends list, with the restriction that friends within a group are aware of each other, but not across groups, and thus do not share content. In figure 1, group 4 friends *Nina, Clint* and *Moira* are *Joe*'s standard friends, and they can get friends recommendations within the groups. For example, *Clint* and *Nina* will get *Moira* as a suggested friend being a mutual friend of *Joe*, but not *Sue* or *Karen*. This is because they are in a different partitions.

6. THE *GREEN*SHIP LANGUAGE

We now propose a declarative language that captures the intended semantics of the model and which allows users to interact with the system. Note that the user interface of *Green*Ship need not be a text interface as proposed. In fact, given the language, a Facebook like graphical user interface is much easier to build.

6.1 Creating and Updating Groups

Users create or modify named groups using the update as statement as follows. In the statement below, *group* is the name of the group and one of {*global, close, standard, exclusive*} is its category. The as option indicates if the membership of this group is visible within and outside this group. The {add|remove|block|report} option allows a list of friends to be added to a newly created or existing group, or removed, blocked or reported from an existing group. In this option, we follow the hierarchy *report*>*block*>*remove* to imply that if a list of friends is reported, then they are also blocked and removed, whereas removal does not constitute automatic reporting. Thus, reporting is a stronger action than blocking and removing.

```
update {global|close|standard|exclusive}
      friends group
   as {visible|invisible}
   perform {add|remove|block|report} friend-list;
```

The semantics of *global, close*, *standard* and *exclusive* follow the description of friendship in the model with the exception that for exclusive friends, visibility does not imply mutual visibility within the group. We like to note that creating partitioned friend groups and exposing them by making them visible defeats the purpose of partitioning and thus reputation defense. But, we prefer to leave that choice to the user by making as invisible and exclusive friends as the default choices (shown as underlined). Also, the as clause is not applicable with perform options *remove, block* or *report.*

Example 6.1. *To create the* green *and* purple *groups making them visible and invisible respectively,* Joe *will use the following two commands.*

```
update close friends green
   as visible
   perform add {Pierre, Max};
update exclusive friends purple
   as invisible
   perform add {Odelia, Pria, Dan};
```

To remove, block and report Odelia*,* Joe *will issue the following command.*

```
update exclusive friends purple
   perform report {Odelia};
```

But to just remove Sue*, he will issue the* update *command*

```
update close friends brown
   perform remove {Sue};
```

which does not result in a reporting. □

6.2 Posting Statuses and Comments

Once friends are added, users are able to share content as status, comment on others' status, comments and folders, and issue "likes" using the post statement below, and subsequently revise or delete them. They are able to limit each post (status, comment or like) to a group that is either *all, friends, exclusive* or *close*, a custom list, or only for the user as *me* using the for option. The *group* keywords have an implicit inclusion hierarchy *all*>*friends*>*exclusive*>*close* that implies an order of visibility. For example, if a post is for *all*, it will be accessible by the entire community including friends in all the groups of the user. However, *friends* makes it accessible to all the groups, and *exclusive* to all the exclusive and close friends. However, posts for *close* friends are not accessible by everyone or friends in non-close group partitions. In other words, *close* friends have the highest visibility right.

[2] Because only close friends are recommended to members in other close friend groups.

```
{post|revise|delete}
      {note|folder|status|comment|reply|like|message}
         content
   for {group|custom|me} friends
   accept {like|comment|share}
   with duration {stream|time|permanent} to
      {stream|time|permanent}
   retract on {remove|block|report}
   retain until {stream|time|permanent};
```

The posting semantics in *Green*Ship and a system such as Facebook is markedly different. In Facebook, for example, there are several different types of friends, but those are designed mostly to prioritize newsfeed, and not to restrict access to friends list or how comments are shared. As noted earlier, in *Green*Ship content (posts, folders, and comments) is shared within the partitions according to the partition membership semantics and visibility rights. We highlight the subtle but very powerful difference between our approach with the rest of the approaches using the following example.

Example 6.2. *In the following post,* Joe *is sharing a status with his close friends (*Joe*'s group 2 and group 3 friends) that will stay on his timeline in perpetuity. He intends to accept likes, comments and allow sharing. But he will not accept any of the comments that will have retention period of less than a year by the commentator. This negotiation will take place fully automatically when the commentator attempts to comment by allowing only compatible ones.*

```
post status "Hello friends!"
   for close friends
   accept {like, comment, share}
   with duration one year to one year
   retract on block
   retain until permanent;
```

Finally, the post will be removed from the newsfeed of all friends Joe *blocks at a later time. Additionally, all the comments these blocked friends made on this post will also be removed (not just made invisible to* Joe *and the blocked friends). Note that such retraction is not possible in Facebook, and making it invisible to the parties involved is not retraction (as others still can see those). As discussed in section 2, with a great deal of effort, accomplishing posting isolation in Facebook is possible using custom lists, but the negotiation as outlined using the* accept *and* with duration *clauses is not possible.* □

6.2.1 Retention Period of Shared Contents

The accept option helps restrict friend reactions, i.e., how they are allowed to react – like, comment or share. The reaction negotiation we have discussed earlier proceeds in two layers. First, a post will have a duration as part of a retraction scheme as spelled out in retain until clause. Therefore, it will disappear after that period regardless taking with it all reactions - comments, like and so on. Retraction also happens for specific comments and likes when someone is unfriended, blocked or reported according to retract on clause. These timelines and provisions are invisible to friends who are viewing the post. Since in *Green*Ship, authors are given the sole copyright, our view is that it does not diminish the rights of friends.

The part that is interesting is the accept with duration clauses that initiates an implicit negotiation between the poster and the commentator, or the commentator and a responder. In the with duration clause, the poster is possibly making a concession and expressing an acceptable duration for which the comment must be made. The only constraint is that a comment must be entered to survive longer than or equal to the duration specified in the with duration clause. For example, if a post is for 10 days, a comment to this post cannot use option *stream* or duration less than 10 days. Thus the duration hierarchy is *permanent*>*time*>*stream* and a match must be in the opposite direction. We adopt *permanent* as default for with duration, and *stream* as a default for retain until options. These options allow maximum duration for a comment/reaction on a post with the shortest lifespan to be most compatible[3]. In example 6.2, *Joe*'s status will stay perpetually (retain until permanent), while he will accept only streaming comments, i.e., less than the life span of the status. Note that streaming means the comments will survive for a single session for all who have access to this comment.

Example 6.3. Joe *posts a message to be visible by all friends in all partitions and to all members of the network (global).*

```
post status "Hello friends!"
   for all friends
   accept {like, comment, share}
   with duration stream to stream
   retract on block
   retain until permanent;
```

Max*'s comment below will not be compatible with the post's reaction duration and thus will not be posted until he changes his entry "1 day" to "stream." Note that even when it has compatible duration, his comment will not be visible to* Joe*'s friends in partitions 1, 2, 4 and his global friends because visibility will disclose* Max*'s friendship with* Joe *to* Nina*, for example, which* Joe *is unwilling to disclose*[4]*. Note that by stating* with duration stream *to* stream*,* Joe *has essentially forced everyone to see* Max*'s comments in streaming mode, and it will not survive beyond a single session.*

```
post comment "Hi Joe. How are you doing?"
   for close friends
   accept {like, comment}
   with duration 1 day to 5 days
   retract on report
   retain until permanent;
```

[3] The two layer durability through the with duration and retain until may seem duplicative or even counter intuitive since we allow the notion of post and comment duration compatibility. One can argue that by matching the post duration and comment duration through retain until, we can decide compatibility without the with duration option. We feel that having the with duration layer, the poster or commentator can choose a different duration than the post's life time, and the minimum level is always decided by the poster of the status or folder. Note that choosing a comment to last longer than the post itself basically becomes ineffective once the post disappears.

[4] A design choice is to allow everyone to see every public post and the community's reaction to it. In theory, since anyone could see the post, commenting on it does not necessarily imply friendship. However, we decided to be conservative and preferred to eliminate the possibility that in *Max*'s comment, any personal disclosure or expression could indicate a friendship, and *Nina* may just guess it right.

Notice that Max*, in his reaction to* Joe*'s post, is also limiting how friends are allowed to react (like and comment), and for what minimum length they must post their comments to his reactions. Once the time limit of a status or comment expires, all the related comments (the subtree) also expire automatically. Thus* Max *is retaining his authorship rights as well, which is compatible with* Joe*'s duration requirement.*

One fine point we would like to mention here is that in traditional systems such as Facebook, ownership of posts stays with the writer and with the original poster of the status or folder. A commentator can and must delete his own comment to make related comments disappear. In Green*Ship, we allow posters and commentators to control reactions to their own status and comments. For example,* Joe *can now respond to* Max*'s comment as follows to make it disappear after* Max *has read it once. Note that it also means that everyone in the green group have streaming access to* Joe*'s response to* Max*. Furthermore, no reaction – comment, like or share – was allowed.*

```
post reply "Fine, thank you!"
   for green friends
   with duration stream to stream
   retract on report
   retain until stream;
```

Also, note that Max*'s comment becomes compatible when he changes his duration to "*stream*" from "*1 day*."* □

6.2.2 Monotonic Restriction of Audience

We also adopt the privacy notion of audience limitation of traditional systems such as Facebook, and thus follow an identical visibility rule albeit with a slight difference. For example, even if *Joe* issued the status in example 6.3 for the world, *Nina* sharing it with the rest of the world will not make it visible to *Joe*'s friends in all the partitions in which she is not a member. Plus, if *Joe* shared the post with only (all) friends, *Nina*'s sharing will only be visible to members of group 4, and no one else. In other words, sharing, commenting and liking are only visible by an ever shrinking group of people, subsets of *Joe*'s intended set and those who can see *Joe*'s content, never to more people than he intended and the model allows.

7. USER RANKING AND REPORTING

In the interest of assisting users weed out potential blind spots, we support two ancillary services. The first one ranks every user on a reputation scale from *highly risky* to *trustworthy*, with *risky, inconclusive*, and *safe* in between. The second service monitors reported users for possible abuse.

We believe, reputed users have a much less likelihood of committing unlawful acts and thus will not engage in cyber abuse or cyber crime because they also have much to lose. If we consider the reputation of a user u as a score ρ_u between the interval 0 and 1, a simple inverse function $(1 - \rho_u)$ can be assumed to be reflective of the risk he poses. These scores can then be converted into the categories such as *highly risky, risky, inconclusive*, *safe* and *trustworthy* using an objective mapping function.

To our knowledge, [22, 23] are among a handful of research efforts that have investigated ways to quantify social network user reputation, taking into account mainly community recommendations and social activities. The quantification suggested in [22] though computable, requires a significant degree of community participation and collaborative filtering. This model parallels enterprize reputation models for e-commerce applications that relies on some form of user feedback. In our view, personal reputation should be based mostly, if not entirely, on user activities and her social presence. Some facts to be taken into consideration could be the length of time they have been friends, the degree or type of social interactions they have had, how reputable a person's friends are, and how much friends value a person's opinions. These measurements should be based on temporal records and non-invasive, behind the scene techniques using users and personal information in an aggregated fashion.

Finally to assist users report, follow up and be protected from potential abuse, it should be possible to alert users if suspected reputation damaging activities are in progress using research similar to [17, 18, 33, 34] to detect text, pics and other materials used as weapons. We believe there are possible ways of developing such a service without violating personal privacy and without potential harmful disclosures. Such a service could be proactive, or on request for specific suspected abusers. We defer a discussion on these services to a separate article considering them as tangential to the issues discussed in this paper.

8. CONCLUSION

We have proposed a novel model for personal reputation management and a declarative language for its implementation leveraging the concept of stratified privacy to help limit reputation damage by malicious friends. We have demonstrated that the added cost to manage multiple categories of friends to achieve stratified privacy more than pays off in increased privacy and convenient posting options compared to Facebook's inadequate and inconvenient list management-based partitioning approach for privacy management.

In section 7, we have briefly described the *Green*Ship system assistance for users to determine online reputation of friends, and active detection of reputation attacks. Our idea of suggesting friends' social reputation was inspired by the work in [22] that attempted to quantify social reputation of any user by analyzing their online behavior and community recommendation. However, we believe that a reputation rating algorithm should also take into account the user's personal behavior and own social risk should she engage in unscrupulous activities to damage others, something that the proposal in [22] did not consider. We also believe that some form of online abuse detection system can be developed along the line of research in [18, 33, 34].

9. REFERENCES

[1] *Bang your friends.* http://www.bangwithfriends.com/. Accessed: November 28, 2016.

[2] *Data Privacy Day.* http://www.microsoft.com/en-us/twc/privacy/data-privacy-day.aspx.

[3] *KnowTheNet.* http://www.knowthenet.org.uk/. Accessed: November 28, 2016.

[4] *One in five young people has suffered online abuse, study finds.* http://tinyurl.com/ntlc7jq. Accessed: November 28, 2016.

[5] *Online Privacy Commissioner of Canada.* http://tinyurl.com/hns2epq. Accessed: November 28, 2016.

[6] *Online Reputation in a Connected World.* http://tinyurl.com/6m2vfeq. Accessed: November 28, 2016.

[7] *Publishing highly personal and embarrassing information about another, even if completely true.* http://tinyurl.com/z7c2tz7. Accessed: November 28, 2016.

[8] *Relationbook.* http://relationbook.me/. Accessed: November 28, 2016.

[9] *Facebook is the worst social network for bullying with 19-year-old BOYS the most common victims.* http://tinyurl.com/ko9hvyg, 2013. Accessed: April 9, 2016.

[10] *Study Reveals that Harassers Prefer Facebook Over Other Networks.* http://tinyurl.com/z4kao7k, 2014. Accessed: April 9, 2016.

[11] *Online abuse affects 3 in 5 Australians: study.* http://tinyurl.com/hat7scx, 2015. Accessed: April 9, 2016.

[12] M. M. Anwar, P. W. L. Fong, X. Yang, and H. J. Hamilton. Visualizing privacy implications of access control policies in social network systems. In *2nd and 4th International Workshop on Data Privacy Management and Autonomous Spontaneous Security, France, September 24-25*, pages 106–120, 2009.

[13] A. Blachnio, A. Przepiórka, E. Balakier, and W. Boruch. Who discloses the most on facebook? *Computers in Human Behavior*, 55:664–667, 2016.

[14] I. Brackenbury and T. Wong. Online profile & reputation perceptions study. Technical report, Microsoft Corporation, 2012.

[15] P. B. Brandtzæg, M. Lüders, and J. H. Skjetne. Too many facebook "friends"? content sharing and sociability versus the need for privacy in social network sites. *Int. J. Hum. Comput. Interaction*, 26(11&12):1006–1030, 2010.

[16] CEOP. *Child exploitation and online protection centre.* http://ceop.police.uk/. Accessed: November 28, 2016.

[17] I. ContentWatch. *Net nanny social.* http://www.netnanny.com/. Accessed: 11/28/2016.

[18] M. de Castro Polastro and P. M. da Silva Eleuterio. A statistical approach for identifying videos of child pornography at crime scenes. In *Seventh International Conference on Availability, Reliability and Security, Czech Republic, August 20-24*, pages 604–612, 2012.

[19] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *WWW*, pages 351–360, 2010.

[20] J. Fox and R. S. Tokunaga. Romantic partner monitoring after breakups: Attachment, dependence, distress, and post-dissolution online surveillance via social networking sites. *Cyberpsy., Behavior, and Soc. Networking*, 18(9):491–498, 2015.

[21] D. Goebert, I. Else, C. Matsu, J. Chung-Do, and J. Y. Chang. The impact of cyberbullying on substance use and mental health in a multiethnic sample. *Maternal and Child Health Journal*, 15(8):1282–1286, November 2011.

[22] D. He, Z. Peng, L. Hong, and Y. Zhang. A social reputation management for web communities. In *WAIM Workshops, Wuhan, China, September 14-16*, pages 167–174, 2011.

[23] J. Y. J. Hong and J. H. Kim. Are social media useful for managing reputation online?: Comparing user interactions online with reputation indicators. In *7th International Conference on Social Computing and Social Media, USA, August 2-7*, pages 207–215, 2015.

[24] D. Isherwood and M. Coetzee. Trustcv: Reputation-based trust for collectivist digital business ecosystems. In *Twelfth Annual International Conference on Privacy, Security and Trust, Toronto, ON, Canada, July 23-24*, pages 420–424, 2014.

[25] C. M. Kokkinos, E. Baltzidis, and D. Xynogala. Prevalence and personality correlates of facebook bullying among university undergraduates. *Computers in Human Behavior*, 55:840–850, 2016.

[26] C. Komisarjevsky. *The Power of Reputation: Strengthen the Asset That Will Make or Break Your Career.* Amacom Books, NY, USA, 2012.

[27] B. J. Litwiller and A. M. Brausch. Cyber bullying and physical bullying in adolescent suicide: The role of violent behavior and substance use. *Journal of Youth Adolescence*, 42:675–684, 2013.

[28] M. Madden and A. Smith. *Reputation Management and Social Media: How people monitor their identity and search for others online.* http://tinyurl.com/jxtkrcs, 2010. Accessed: November 28, 2016.

[29] T. Nakano, T. Suda, Y. Okaie, and M. J. Moore. Analysis of cyber aggression and cyber-bullying in social networking. In *IEEE ICSC, Laguna Hills, CA, USA, February 4-6*, pages 337–341, 2016.

[30] NCPC. *Stop cyberbullying before it starts.* http://tinyurl.com/bz7axqb. Accessed: November 28, 2016.

[31] M. O'Neill. *Diesel cam brings facebook to the fitting room.* http://tinyurl.com/zv6ynl3, May 2010. Accessed: July 28, 2016.

[32] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21*, 2011.

[33] S. P. Ros, Á. P. Canelles, M. G. Pérez, F. G. Mármol, and G. M. Pérez. Chasing offensive conduct in social networks: A reputation-based practical approach for frisber. *ACM Trans. Internet Techn.*, 15(4):15, 2015.

[34] K. V. Royen, K. Poels, W. Daelemans, and H. Vandebosch. Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability. *Telematics and Informatics*, 32(1):89–97, 2015.

[35] A. M. Schenk and W. J. Fremouw. Prevalence, psychological impact, and coping of cyberbully victims among college students. *Journal of School Violence*, 11:21–37, 2012.

[36] C. Seebach, R. Beck, and O. Denisova. Analyzing social media for corporate reputation management: How firms can improve business agility. *IJBIR*, 4(3):50–66, 2013.

[37] R. D. Wolf, K. Willaert, and J. Pierson. Managing privacy boundaries together: Exploring individual and group privacy management strategies in facebook. *Computers in Human Behavior*, 35:444–454, 2014.