DOI:
10.1109/MIC.2017.4180831

Document Version
Peer reviewed version

PACMAN: Personal Agent for Access Control in Social Media

Gaurav Misra and Jose M. Such

Abstract— Given the plethora of interactions that social media users engage in, appropriately controlling access to their information becomes a challenging task for them. Selecting the appropriate audience, even from within their own friend network, can be fraught with difficulties especially considering the dynamic nature of the medium. In this paper, we propose PACMAN, a personal assistant agent that recommends personalized access control decisions based on the social context of any information disclosure by incorporating communities generated from the network structure of the user and utilizing information in their profiles, in addition to the particular content to be shared. We show in this paper that PACMAN provides highly accurate recommendations while minimizing intrusiveness via a thorough empirical evaluation using a collected dataset of actual access control decisions.

Index Terms—Social Media, Machine Learning, Access Control

1 Introduction

SOCIAL media has become synonymous with communication in daily life for most of us. On Facebook alone, over 1 billion users share over 300 petabytes of personal information daily. Social media users interact with people representing various facets of their life such as work, family, education, etc. In such a scenario, it is essential for them to make informed access control decisions to preserve the "contextual integrity" of their information. Any user who discloses information on social media has a notion of the "intended recipients" and the context in which they would view that information and hence preservation of "contextual integrity" is essential to avoid a privacy breach [1]. Unfortunately, the privacy controls afforded to users by social media sites make it burdensome to selectively share content within their friend network which results in a situation where they end up sharing their information with "unintended recipients" [2]. The mainstream social media sites such as Facebook and Google+ have taken steps to mitigate this by assisting users in managing their friend networks by creating Lists and Circles respectively. However, recent research findings suggest that hardly any users employ these features when making access control decisions, arguably due to the effort this requires from them [3].

Social media users can be assisted by recommendation systems which can guide them towards the appropriate access control decisions. It is well established that different users exhibit different access control behavior and often have differring privacy preferences. Therefore, it is essential that a recommendation system forms the core of a personal agent which can provide personalized recommendations to individual users. In recent times, we have seen personal agents being proposed to provide assistance to users in various social media issues such as ascertaining contexts of disclosure [4], detecting privacy violations when they happen [5] and negotiating multi-party privacy conflicts [6]. However, to the best of our knowledge, there is an absence of a personal agent which recommends personalized access control decisions to users to minimize the burden of expressing their individual sharing preferences. In this paper, we present PACMAN, a personal agent which provides a novel approach to learning access control decisions by combining social relationships and information about the content. The building blocks of PACMAN are identified by conducting detailed empirical evaluations that result in a highly accurate mechanism using a minimal set of appropriate attributes. Our results show that PACMAN produces an average accuracy of 91.8% (sd=6.5%, median=94.1%) across all users. We find that PACMAN works best for users who are more *static* in terms of the number of friends they grant access to.

2 PACMAN

Figure 1 shows the information that PACMAN uses to produce an access control recommendation ("allow" or "deny"). For social media users, the social context of information disclosure is considered essential to enable the formulation of access control policies in a way which preserves the "contextual integrity" of the information [7]. The social context can be derived from information which facilitate the definition of social relationships on these media. These interpersonal relationships can be defined in terms of *relationship types*, often denoted by communities [8] and *relationship strength* or closeness which is represented by similarity of profile attributes [9], [10]. In addition to relationships, the content itself is an integral part of the context of the disclosure and plays an important role in the formulation of a desired access control policy.

- *Gaurav Misra is a PhD student at Lancaster University, Lancaster, UK. E-mail: g.misra@lancaster.ac.uk*
- *Dr. Jose M. Such is a Senior Lecturer at King's College London, UK. E-mail: jose.such@kcl.ac.uk*

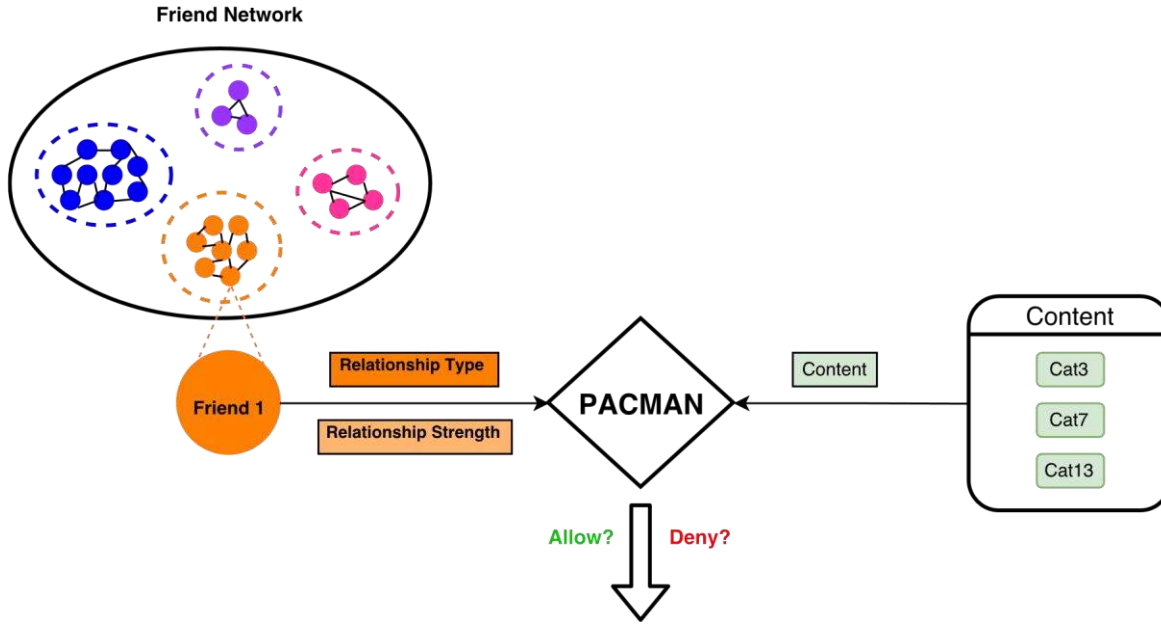

Fig. 1. Various components and inputs of PACMAN

Therefore, the information about the content being shared (text, photos, etc.) should also be used to learn access control decisions [11].

2.1 Relationship Type

Social media users have various types of interpersonal relationships (eg: friends, colleagues, family, etc.) with the people they interact with on the network. These can often be represented by partitioning one's network into groups or communities. These partitions can then be leveraged by any access control mechanism such that the user may be asked to make access control decisions with respect to one or some members in a particular "community" (created by the algorithm), and then implement that decision for the other members in that community [12]. PACMAN uses network based community detection and requires the friend network of the user.

2.2 Relationship Strength

The interpersonal relationships between social media users can also be defined in terms of strength (or closeness). This is generally estimated by measuring similarity between profiles of individuals. There have been several proposed approaches in literature which suggest appropriate methods of estimating tie-strength or closeness such that it can be used to assist users in making informed access control decisions [13]. However, they all have the same limitations: 1) The information required from the profiles may not be *easy to fetch and process* and makes it difficult to provide users with real time assistance on a dynamic medium like social media; 2) Some profile attributes may often be *missing* as users often refrain from populating many fields on their social media profiles [9] and 3) Accessing certain types of personal information from the users' profiles may be *intrusive* and hence counter productive for a privacy preserving mechanism. In our previous work [14], we performed a systematic analysis of all profile attributes available in social media profiles to select the minimal subset most suitable for predicting access control decisions with maximum possible accuracy. The analysis led to the identification of *Total Friends* (total size of a user's friend network) and *Mutual Friends* (number of shared friends or contacts with the user) as the most appropriate profile attributes to enable prediction of access control decisions, while overcoming the discussed challenges. Therefore, PACMAN uses these two attributes to account for relationship strength between a user and each of their friends.

2.3 Content

The information about the content being shared can also be used to enhance access control mechanisms. The information about the content can be automatically mined and used to classify the content which can then be leveraged to inform access control decisions [15]. Different methods can be used to create attributes depending on the nature of the content (for eg: natural language processing techniques can be used for text and image processing can be used for photos). However, such analysis is still far from being completely automated in terms of accuracy to represent the user's perception about the content. A particular method of mitigating this is by asking users to provide metadata in the form of "tags" while sharing the content. Previous research shows that such tags can also be used to create access control policies and that they are minimally disruptive for the user [16]. PACMAN is agnostic to the type of content being shared and hence different methods of obtaining information about the content can be implemented. If automatic analysis of content is implemented in PACMAN, it can operate completely without any user input as the other attributes, relationship type and relationship strength, are analyzed automatically by PACMAN.

3 Evaluations

To evaluate whether and to what extent access control decisions made by social media users can be learned by PACMAN, we conducted a user study in order to obtain ground truth access control decisions to use for learning which is the standard way of evaluating automated access control mechanisms in the literature.

3.1 Experiment

We created an application using Facebook Query Language (FQL) and the Facebook Graph API for participants to make access control decisions while disclosing 10 photos. Five of these photos were randomly downloaded from their Facebook profiles, and the participants were asked to select and bring 5 other photos which they had not yet uploaded on Facebook in order to avoid a scenario where a user makes access control decisions for all photos during the study for which they had already received comments and likes before, as that may have influenced their decisions. The participants were advised to bring photos which they considered to be personal (either included them or a family member) or considered sensitive so that they had a privacy implication. The different stages were:

1) The participants logged into the application using their Facebook credentials. They were then alerted about the data that would be accessed and asked for explicit permissions before moving on.

2) The participants were shown 10 photos sequentially on the screen, each on an individual page. They were asked to select categories for the photos from a predefined list of 15 popular Flickr categories, and make access control decisions for each photo. The friend list was shown alphabetically to the participants and they were instructed to select each and every friend that they would want to grant access to the photo. They were explicitly informed that any friend who was not selected would be denied access to the photo.

3) Once participants made the access control decisions and selected the categories for all 10 photos, their selections, friend lists and *Total Friends* and *Mutual Friends* profile attributes of all their friends were stored.

3.2 Participants

This research experiment was conducted at Lancaster University after being approved by the Research Ethics Committee of the university. Participants were recruited primarily from among the staff and students of the university. Additionally, we invited some participants who were external to the university through personal communication channels such as email, social networks, etc. All participants were compensated with £10 for their involvement in the study.

We applied the typical pre- and post-experiment checks to maximize data quality. In particular, before the experiment we screened participants and everyone who had a Facebook account and had uploaded at least 10 photos before the study was eligible to participate. After an initial registration phase, 31 participants were selected who took part. After completion of the user study, we checked all responses to make sure participants had correctly completed the experiment, finding 5 participants who did not (4 had randomly selected lists of alphabetically sorted friends, 1 had selected one single but different friend for each photo). The remaining 26 participants were considered for the analyses, including 15 males (57.7%) and 11 females (42.3%). The average age of the participants was 29 years (s.d = 6) and the average size of network was 265 friends (s.d = 121). The total number of access control decisions made by the 26 participants during the experiment, and hence the size of the ground truth dataset, was 67,660.

3.3 Implementation of PACMAN

The design of PACMAN described earlier in this paper was implemented using various building blocks to represent the different components shown in Figure 1. The information required from the users' Facebook profiles and their access control decisions were obtained from the user study as described.

To represent *Relationship Type*, PACMAN uses community membership. In our previous work [17], we evaluated 8 wellknown network based community detection algorithms for a goodness of fit with access control decisions made by social media users. Our analysis found *Clique Percolation Method (CPM)* to be the most suitable community detection algorithm in an access control scenario and CPM membership is used to represent *Relationship Type* in this implementation of PACMAN. The friend network of each user obtained during the user study was used as input to the CPM algorithm, implemented using the iGraph library, to create communities. Each of a user's friends was assigned a community membership which was denoted using a binary vector, with dimension equal to the total communities of the user, to represent their relationship type in PACMAN. For this implementation, we used non-overlapping CPM communities such that each of the users' friends belonged to exactly one particular community.

For *Relationship Strength*, the *Total Friends* and *Mutual Friends* attributes were directly fetched from the users' profiles during the study and used as an input to the PACMAN mechanism.

As mentioned earlier, the design of PACMAN is agnostic to the type of content being shared as well as the method used to obtain information about the content. In this particular implementation, we used manual selection of photo categories in the form of "tags" to represent the information about content. This was done as it provided us with the user's perspective about the content in a comparatively less intrusive way [16]. The users during the study were given an opportunity to select categories for the photos in the form of tags as mentioned earlier. While it was not mandatory to select categories for each photo, we found only 4 out of the 260 photos (10 per user) which

were not categorized. The average number of categories selected per photo was found to be 2.2. The content information was represented with a binary vector having a dimension of 15 (the total number of categories) representing whether each category was selected or not. Thus, a photo which was not categorized would be represented as all zeroes.

For evaluating PACMAN's performance, Weka was integrated into PACMAN to create and run the classifier using 10-fold cross validation to calculate accuracy of prediction produced for each individual user. There were 67,660 instances in total corresponding to all the access control decisions in the ground truth dataset. The attributes consisted of the CPM membership vector, total and mutual friends as well as the content vector representing the photo categories. In 10-fold cross validation, the entire dataset is randomly divided into 10 subsets, each of which are then used as training data (while leaving the other 9 as test sets) for each iteration. This process is repeated 10 times such that each subset gets to be the training set and the average error across all 10 iterations is considered as the final value. We performed 10-fold cross validation using the in-built function present in Weka which automatically divides the dataset into 10 random subsets. To the best of our knowledge, this is the most rigorous and systematic method of evaluating a classifier as it rules out the possible bias associated with division of a dataset into training and test sets.

PACMAN can work with any machine learning algorithm and for the evaluation, we tried Naive Bayes classification algorithm, Support Vector Machines (SVM) as well as Random Forest but found that Random Forest produced the best results and have only reported those in this paper due to lack of space.

3.4 Estimating User Effort

PACMAN recommends "allow" or "deny" access control decisions to the user corresponding to each member in their friend network. For PACMAN, both classes, "allow" and "deny" are of equal importance, as users would spend time and effort in correcting the erroneous recommendations made by PACMAN. In such a scenario, accuracy is appropriate as other metrics focus on giving more importance to one of the classes [18], e.g: when a program is to be classified as malware or not, positive classification is prioritized. To calculate accuracy, we use the access control decisions made by the users for all the 10 photos during the user study as the ground truth. In particular, for a user having F total friends, the accuracy of PACMAN can be calculated as a percentage of the total recommendations that are correct:

$$Accuracy = ((F - Errors)/F) \quad (1)$$

The *Errors* include both "allow" and "deny" errors: An *Allow Error* occurs when PACMAN recommends a "deny" decision to the user when it actually should have been "allow". These errors are essentially "False Negative" (FN) recommendations and result in a "deny to allow" change being made by the user.

A *Deny Error* occurs when PACMAN recommends an "allow" decision to the user when it actually should have been "deny". These errors are "False Positive" FP recommendations and result in an "allow to deny" change by the user.

$$Errors = FN + FP \quad (2)$$

We show the ratio of both types of errors for each user to provide a more precise picture of the performance of PACMAN regarding each type of error.

In addition to reporting the accuracy of the recommendations made by PACMAN, we also show the area under ROC curve (AUC) to give an idea of the quality of recommendations made by PACMAN.

4 RESULTS

4.1 Overall Accuracy

Figure 2 shows the accuracy of recommendations produced by PACMAN for each of the 26 users. It also shows the ratio of incorrect recommendations which were *Allow Errors* and *Deny Errors*. We can see from the figure that PACMAN produces highly accurate recommendations for almost all the users. The average accuracy across 26 users was found to be 91.8% (sd=6.5%, min=81.4%, max=99.7%). We find that almost all users have similar amounts of Allow Errors (mean=5.4%, sd=3.9%) and Deny Errors (mean=2.8%, sd=3.5%). This suggests that PACMAN does not discriminate between the two classes and that recommendation errors are fairly equal. The average area under ROC curve (AUC) was 0.845 (sd=0.097) which shows PACMAN produces good quality recommendations.

4.2 Clustering Users

In order to enhance our understanding of the strengths and weaknesses of PACMAN, we wanted to examine the factors which may distinguish the users for whom it produces high accuracy as compared to the ones with comparatively lower accuracy. We used two-step clustering, using overall accuracy as the clustering variable, to obtain the two clusters of users as described in Table I.

We find a cluster of 17 users for whom PACMAN produces very high accuracy (mean=96.1%, stdev=2.9%). These users have a comparatively more static access control behavior, with lower average and standard deviation for audience sizes (across 10 photos), and smaller number of communities. The 9 users in the other cluster were found to have comparatively lower but still decent accuracy (mean=83.8%, stdev=1.9%). It is noticeable that they have greater variation in their access control behavior with higher average and standard deviation for audience sizes and number of communities. The table also shows that both clusters have similarly high AUC values which suggests that PACMAN produces good quality recommendations for all users.

We also calculated the correlation coefficients with respect to accuracy and the access control behavior of users. These coefficients are shown in Table II. The correlations confirm the hypothesis that users who have larger aver

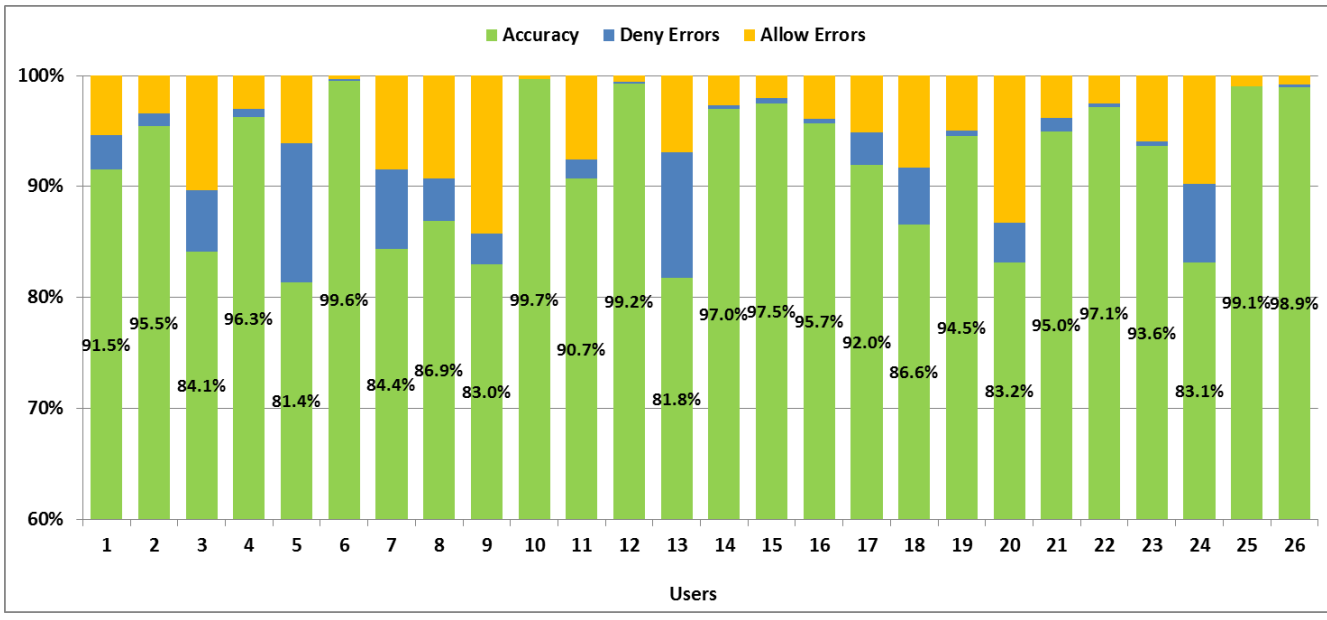

Fig. 2. Accuracy and ratio of changes required to recommendations made by PACMAN for all 26 users

TABLE 1
PACMAN accuracy and access control behavior of users in both clusters

Cluster	Users	Statistic	Average Audience	Std. Dev. Audience	Comms. Used	Accuracy	Allow Errors	Deny Errors	AUC	Relative Information Gain*		
										Type	Strength	Content
Higher Accuracy	17	Average	15.06	19.65	5.18	96.1%	3.1%	0.8%	0.848	0.170	0.490	0.339
		Stdev	14.30	22.34	3.88	2.9%	2.2%	0.9%	0.118	0.290	0.409	0.383
Lower Accuracy	9	Average	59.87	57.46	10.33	83.8%	9.6%	6.6%	0.838	0.155	0.336	0.509
		Stdev	20.61	17.82	7.60	1.9%	2.7%	3.4%	0.052	0.162	0.317	0.317
Overall	26	Average	30.57	32.74	6.96	91.8%	5.4%	2.8%	0.845	0.165	0.437	0.398
		Stdev	27.19	27.52	5.86	6.5%	3.9%	3.5%	0.097	0.249	0.381	0.365

*The difference in relative information gain values was not found to be statistically significant between the clusters. All other differences were found to be statistically significant at the 99% confidence interval using the Mann-Whitney Test.
Average and Stdev values show aggregate statistics across all users of a particular cluster.

age audiences and larger variations in their selections are more likely to have higher errors (both *Allow Errors* and *Deny Errors*) and a comparatively lower accuracy as a result.

We did not find any significant correlations in terms of the personal characteristics of the users such as gender, age, number of photos uploaded (amount of activity on Facebook) or size of the friend network. No significant trends could be observed with respect to the category or source (Facebook or USB) of photos in terms of accuracy of PACMAN prediction. This suggests that PACMAN would work for all categories of photos and whether they had been previously been uploaded on social media does not have an effect on its performance.

4.3 Contribution of Types of Attributes

We wanted to examine whether all three types of attributes were required and were contributing to the performance of PACMAN or whether one or more were redundant and could be avoided without compromising the performance. We calculated the relative information gain for each type of attribute as a ratio of the total information gain in order to compare the contribution for each individual user.

TABLE 2
Pearson correlation of accuracy and contribution of components with access control behavior

	Average Audience	Std. Deviation Audience	Communities Used
Allow Errors	0.660**	0.576**	0.360
Deny Errors	0.896**	0.800**	0.636**
Accuracy	-0.880**	-0.777**	-0.558**
AUC	0.198	0.362	0.130
Rel. Type Gain	-0.149	-0.194	0.036
Rel. Strength Gain	-0.312	-0.402*	0.209
Content Gain	0.428*	0.553**	0.194

**Correlation is significant at the 99% confidence level
*Correlation is significant at the 95% confidence level

The aggregated values for for all 26 users as well as both clusters of users are shown in Table I. The numbers suggest that all components contribute to the performance of PACMAN while *Relationship Strength* seems to contribute the most for the average user. The difference between the clusters was not found to be statistically significant. Nevertheless, the numbers suggest that PACMAN relies more on the *Content* for users who show

greater variation in their access control behavior. This notion is also supported by the correlation coefficients in Table II where we find that PACMAN relies more on *Content* for users who select larger audiences and have greater variation. Therefore, it is plausible that the PACMAN accuracy for such users would improve by training with more photos for each type of photo content.

5 Related Work

There have been many previous works in the area of predicting and recommending access control decisions to social media users. Many of these works use different types of attributes to enable prediction of access control decisions. Some approaches advocate the user of community membership [8] while others rely on profile information [9], [19] to recommend access control decisions. The information about the content can also be used to determine the appropriate audience to be selected [15]. PACMAN advances the state of art by using a conjunction of relationship based attributes, communities and profile attributes, to represent the "who" and the information about the content, the "what", to represent the social context of disclosure.

6 Conclusions and Future Enhancements

In this paper, we presented PACMAN, which leverages information about interpersonal relationships between individuals on social networks and combines it with user supplied information about the content to recommend access control decisions. Our evaluations show that PACMAN produces highly accurate access control recommendations and all three components of PACMAN are important, and each individual component has varying importance for different users. PACMAN is found to rely on the information about the content more for users who have greater variation in their access control behavior.

Having considered only network based community detection for representing relationship type in PACMAN, we can consider social circles, based on contextual information, beyond social media profiles, such as co-location [20], as a possible future enhancement. Sensors on mobile devices can be used to identify contacts in the same location which could be used as an attribute [20]. Looking at the reliance of PACMAN on content for the users with greater variation in access control behavior, other methods of extracting information about content such as the physical properties of the photos themselves [11] could be considered to observe if it enhances the accuracy for such users. This would enable PACMAN to function without any user input and make it work in a scenario where a social network is a network of agents which make access control decisions based on automatic analysis of the attributes. Finally, PACMAN focuses on learning individual preferences, which may also be used as input to other tools that recommend access control decisions for multi-user scenarios [21].

References

[1] H. Nissenbaum, "Privacy as contextual integrity," *Washington Law Review*, vol. 79, p. 119, 2004.

[2] S. Egelman, A. Oates, and S. Krishnamurthi, "Oops, i did it again: mitigating repeated access control errors on facebook," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2011, pp. 2295–2304.

[3] P. Wisniewski, B. P. Knijnenburg, and H. Richter Lipford, "Profiling facebook users privacy behaviors," in *SOUPS2014 Workshop on Privacy Personas and Segmentation*, 2014.

[4] N. Criado and J. M. Such, "Implicit contextual integrity in online social networks," *Information Sciences*, vol. 325, pp. 48–69, 2015.

[5] O. Kafali, A. Gunay, and P. Yolum, "Protoss: A run time tool for detecting privacy violations in online social networks," in *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*. IEEE, 2012, pp. 429–433.

[6] J. M. Such and M. Rovatsos, "Privacy policy negotiation in social media," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 11, no. 1, p. 4, 2016.

[7] G. Misra and J. M. Such, "How socially aware are social media privacy controls?" *Computer*, vol. 49, no. 3, pp. 96–99, 2016.

[8] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *Proc. of the 19th international conference on World wide web*. ACM, 2010, pp. 351–360.

[9] S. Amershi, J. Fogarty, and D. Weld, "Regroup: Interactive machine learning for on-demand group creation in social networks," in *Proc. of the SIGCHI*. ACM, 2012, pp. 21–30.

[10] J. J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks." in *NIPS*, vol. 272, 2012, pp. 548–556.

[11] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede, "Privacy policy inference of user-uploaded images on content sharing sites," *IEEE transactions on knowledge and data engineering*, vol. 27, no. 1, pp. 193– 206, 2015.

[12] G. P. Cheek and M. Shehab, "Human effects of enhanced privacy management models," *Dependable and Secure Computing, IEEE Transactions on*, vol. 11, no. 2, pp. 142–154, 2014.

[13] R. L. Fogues, 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*, pp. 1–13, 2013.

[14] G. Misra, J. M. Such, and H. Balogun, "Improve - Identifying Minimal PROfile VEctors for similarity based access control," in *2016 IEEE Trustcom/BigDataSE/ISPA*, Aug 2016, pp. 868–875.

[15] 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.

[16] C.-m. A. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing access control to online photo albums based on tags and linked data." in *AAAI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0*, 2009, pp. 9–14.

[17] G. Misra, J. M. Such, and H. Balogun, "Non-sharing communities? an empirical study of community detection for access control decisions," in *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE, 2016, pp. 49–56.

[18] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.

[19] C. Akcora, B. Carminati, and E. Ferrari, "Privacy in social networks: How risky is your social graph?" in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 2012, pp. 9–19.

[20] P. Murukannaiah and M. Singh, "Platys social: Relating shared places and private social circles," *IEEE Internet Computing*, vol. 16, no. 3, pp. 53–59, 2012.

[21] R. L. Fogues, P. Murukannaiah, J. M. Such, and M. P. Singh, "Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making," *ACM Transactions on Computer-Human Interaction*, p. In press., 2017.

Gaurav Misra is in the final stages of his PhD research at Lancaster University in UK. His doctoral research focuses on creating solutions to access control problems faced by social media users. He will soon start working as a postdoctoral research fellow at University of New South Wales (UNSW) in Canberra, Australia, where he will be working on tackling problems emanating from the social aspects of security and privacy.

Jose M. Such is Senior Lecturer (Associate Professor) in Computer Science at King's College London. His research interests are at the intersection between cyber security, artificial intelligence, and human-computer interaction; with a strong focus on privacy, intelligent access control, and co-owned data in socio-technical and cyber-physical systems.