Enhancing Privacy in Online Social Communities: Can Trust Help Mitigate Privacy Risks?

Venkata Swamy Martha[1], Nitin Agarwal[2], and Srini Ramaswamy[3]

[1] @WalmartLabs, California, USA
vmartha@walmartlabs.com
[2] University of Arkansas at Little Rock, AR, USA
nxagarwal@ualr.edu
[3] ABB Corporate Research, Bangalore, India
srini@ieee.org

Abstract. The context based privacy model (CBPM) has proved to be successful in strengthening privacy specifications in social media. It allows users to define their own contexts and specify fine-grained policies. Collective-CBPM learns the user policies from community. Our experiments on a sample collection of Facebook data demonstrated the models feasibility in real time systems. These experiments however, did not capture all of the user scenarios; in this paper we simulate users for all possible user scenarios in a social network. We operationalize the C-CBPM model and study its functional behavior. We conduct experiments on a simulated environment. Our results demonstrate that even the most conservative user never incurs risk greater than 20%. Moreover, the risk diminishes to 0 as the trust increases between donors and adopters. The model poses absolutely no risk to other liberal or semi-liberal users.

Keywords: context based privacy model, collective-CBPM, access control, trust, collective intelligence, social media.

1 Introduction

Though there exist privacy concerns, users are motivated to stay with online societies and participate in its user-friendly environment. Additionally, social media poses a challenging problem, i.e., privacy of Personally Identifiable Information (PII). Privacy aware systems address it by deciding "what" information to share. There is a significant body of work in studying security and privacy. However, there is a significant need for improving personal privacy especially in social media eco-systems.

In our previous work, we implemented a context based privacy model(CBPM)[1] that allows users to define their own contexts and specify fine-grained policies. Borrowing concepts from collective intelligence [2], the model further extended to recommend privacy policies for a user[3]. We performed experiments on a sample of real data collected from Facebook and presented the results in [4]. These experiments however, did not capture all of the user scenarios; in this paper we simulate users for various possible user scenarios in a social network.

R. Natarajan (Ed.): ICDCIT 2014, LNCS 8337, pp. 293–298, 2014.

Section 2 describes CBPM to illustrate C-CBPM model, Section 3 discusses the simulation details, Section 4 presents results and discussio, followed by Section 5 drawing conclusions from the results and ideas for further research.

2 Background

Context Based Privacy Model (CBPM): A context is defined as an abstract state of a subject and rich in expressing status of an entity compared to user or role identity. A privacy policy is defined as a rule to/not to share a data element from a given data set in a given context. The privacy policies for all the data sets and for all the contexts makes up a matrix called CBPM matrix. An entry in a CBPM matrix can be represented as:

$$(C_i, D_j) = 0 \ or \ 1 \tag{1}$$

where, C_i is a context, and D_j is a data set.

Collective- Context Based Privacy Model (C-CBPM): Donors, who are ready to donate, provide their CBPM matrices in C-CBPM donation pool. The donor could select parts (rows and columns) of his/her matrix to make a donation. Adoption of a matrix from C-CBPM involves three phases. 1: Requesting Donations, 2: Processing Requests, 3. Aggregation of donations. The aggregation matrix for an adopter 'X' is obtained using the following formula.

$$\cancel{M}_X = \cancel{M}_X +_{OR} [t(A,X)\cancel{M}_A +_{OR} t(B,X)\cancel{M}_B +_{OR} t(C,X)\cancel{M}_C]_\mu \tag{2}$$

Where 'M_A' is a donation matrix from user 'A'.

'$\cancel{M}_A$' is the masked matrix for the donation 'M_A' and is defined as,

$$\cancel{M}_A(i,j) = M_A(i,j) \ if \ adopter \ accepts \ donation \ for \ the \ context \ `i' \ and \ dataset \ `j'$$
$$= 0 \qquad otherwise$$

The matrix binarization operation '$[]_\mu$' on a matrix is defined as,

$$\begin{aligned}[t(A,X)M_A(i,j)]_\mu &= 1 \quad if \ \ t(A,X) * M_A(i,j) > \mu, \\ &= 0 \quad otherwise\end{aligned}$$

'μ' is a binarization threshold. The threshold value ranges from '0' to '1'. And the matrix aggregation operator '$+_{OR}$' is defined as,

$$\begin{aligned}(M_A +_{OR} M_B)(i,j) &= 0 \quad if \ M_A(i,j) = M_B(i,j) = 0, \\ &= 1 \quad otherwise\end{aligned}$$

It is trivial to show the $+_{OR}$ addition operation is commutative, associative, transitive, and closed. Each donation is associated with a trust value denoted as t(A,X). Trust values always lie in the range of [0,1]. Eq 2 gives us derived matrix for user 'X' from available donations. More detailed discussion of the C-CBPM model was presented in [3].

3 Simulation

Due to operational challenges with integration of the CCBPM in a real world social network (e.g., Facebook) at a large scale, in this section we simulate a social network of users with a synthetic privacy policies.

Simulation Parameters. There exists a set of users called user pool and corresponding CBPM matrices. Strength of the privacy policy of a user, is measured through “Privacy Index”, which is defined as,

“Privacy Index (PI): Fraction of zeros in the CBPM matrix”[1]

Hence, a higher value of PI indicates the conservativeness of a user. Consider a user ‘A’ from the user pool as an adopter (learning privacy policies from community) and rest of the users are donors. Though the adopter ‘A’ embodies a CBPM matrix, we apply C-CBPM for the user ‘A’ to generate suggestion matrix derived from donations. We call the actual matrix of the adopter ‘A’ as “Expected Matrix (EM)” and the recommended matrix from C-CBPM model as “Observed Matrix (OM)”. This strategy helps in two ways One, every user has a CBPM matrix (EM) which can serve as the ground truth for validation. Second, this allows us to quantify trust between adopter ‘A’ and donors. Though similarity may not always imply trust or vice versa, however research has shown strong evidence to counter this intuition [5]. We used EM of the adopter ‘A’ and CBPM matrix of a donor to calculate trust between the adopter ‘A’ and the donor, as defined below,

“Trust (t): Fraction of common entries in the matrix”

It must be noted that the model is adaptable to explicit trust scores. Trust values always lie in the range of [0,1]. Inducing acquired knowledge from trust scores, we classified the donors into several trust domains.

“Trust Domain (TD)[a,b]: Set of donors having trust scores with an adopter between ‘a’ and ‘b’ ”’

We considered Trust Domain interval as 0.1 and each Trust Domain is represented by its upper limit. For example, TD=0.7 represents Trust Domain (0.6, 0.7][2] . To quantify a donation, we introduce a parameter called “Fraction of Donation (FoD)”. FoD lies in [0,1]. The binarization parameter ‘μ’ in equ. 3, we call it threshold, ranges from ‘0’ to ‘1’. We considered every ‘μ’ in (0,1] with an interval of ‘0.1. For an adopter ‘A’, C-CBPM exploits available donor matrices for a given FoD, TD, μ and generates “Observed Matrix (OM)”. To address diverse real time scenarios, we apply C-CBPM for every FoD in (0,1], TD in (0.1,1], μ in (0,1] with an interval of 0.1.

[1] It is extremely challenging to quantify privacy as different attributes could have very different privacy sensitivity. The proposed Privacy Index measure is a very basic attempt toward this direction.

[2] The notation (a,b] denotes that the interval is open at a and closed at b. In other words, for any x, a¡xb.

Evaluation Metrics. Accuracy shows precision of the model for given parameters.

$$Accuracy = \frac{\sum_{i,j} \left(EM\left(i,j\right) \equiv OM\left(i,j\right)\right)}{\sum_{i,j} 1} \tag{3}$$

Here the operator '≡' refers to boolean equivalence and is written as "$a \equiv b = \neg(a +_{XOR} b)$".

Next evaluation metric is "Risk Factor" that assesses the risk posed by a wrong recommendation. In other words, Risk Factor denotes the number of times an unshared data element is recommended to to share. This amounts to the risk incurred from the policies suggested by C-CBPM. Mathematically, Risk Factor is defined as,

$$RiskFactorRF = \frac{\sum_{i,j} \left(\neg\left(EM\left(i,j\right) \wedge OM\left(i,j\right)\right)\right)}{\sum_{i,j} 1} \tag{4}$$

Likewise, we also calculate strength of privacy policy in OM by calculating Privacy Index (PI). The PI of observed matrix is represented by PI_{OM}.

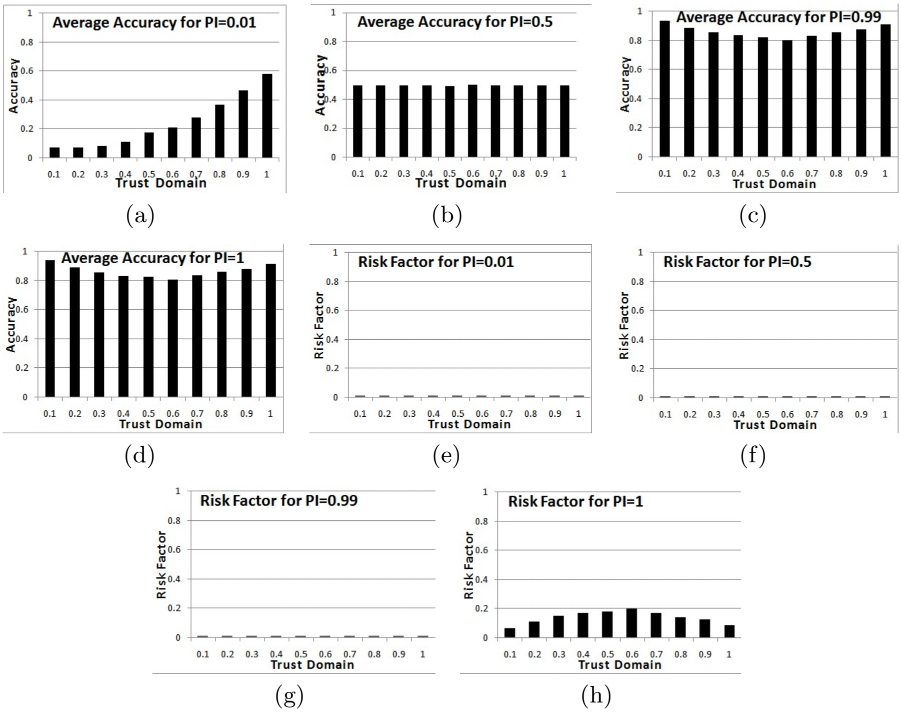
3.1 Experimental Setup

The experiment is carried out in three phases: 1. Synthetic donation preparation, 2. Applying C-CBPM, 3. Evaluating the observations.

- Phase 1. Synthetic donation preparation: We can generate a matrix of size 1000x1000 with desired PI (desired number of zeros in a matrix) using Java's random number generation method. Here, we created a user pool of 100 users with PI values from 0 to 1 with an interval of 0.01.
- Phase 2. Applying C-CBPM: A user from the user pool is selected to name it adopter 'A' and its matrix is expected matrix (EM). Except the adopter, all the other users in user pool are donors. Therefore, we have 99 donors. With FoD=0, TD=0.1, and μ=0, the C-CBPM model generates OM for the adopter 'A'. The OM is evaluated with EM using the given metrics. This step (C-CBPM) is iterated by incrementing FoD, TD, and μ by 0.1 until 1.0. The whole phase is repeated by selecting other users in user pool as adopter, individually.
- Phase 3. Evaluation: The EM and OM are compared using "Accuracy", RF and PI metrics.

3.2 Simulation Results

We tabulated the results from each iteration. Each run in simulation is a row in the table. There are 100,000 rows in table (100 users * 10 TD * 10 FoD * 10). We then summarized the table for each user, i.e., for each EM having PI in (0,1]. For an EM and TD, we estimate average, minimum, and maximum of evaluation metrics i.e. Accuracy (A) and Risk Factor (RF). To demonstrate the models capability we chose users with $PI_{EM} = 1$, 0.99, 0.5 and 0.01 (from 'share nothing' to 'share all') and presented the plots from the table in Figure 1.



(a) (b) (c)

(d) (e) (f)

(g) (h)

Fig. 1. Plots depicting behavior of C-CBPM for adopters with PI=0.01,0.5,0.99,1[3]

4 Discussion

Due to space constraints, we present detailed analysis for users with extreme characteristics, i.e., PI = 0, 0.5, and 1.

- Conservative user (PI_{EM}=1, i.e., 'share nothing'): From figure 1(d), the model always achieves high accuracy with the lowest value of 80%. Regardless of the trust values with the donors, a conservative user following the C-CBPM model will never incur risk more than 20% as shown in figure 1(h). Further the risk can also be reduced by adopting donations from highly trusted donors and less trusted donors.
- Semi-liberal user (PI_{EM}=0.5): The average accuracy for the semi-liberal adopter (figure 1(b)), i.e., PI_{EM}=0.5, is 50%. The model promises at least 50% accuracy irrespective of the other parameters. More importantly, figure 1(f) shows that the risk of adopting the recommended CBPM matrix is always zero to the adopter. Since the risk factor is zero, not even a single private element is made public.
- Liberal user (PI_{EM}=0.01, i.e., 'share all'): Here in all the instances, the adopter gets donations from more conservative user. This makes the model less accurate (figure 1(a)) compared to figures 1(b), 1(c) and 1(d). Though

the accuracy of liberal adopter increases with the trust of the donors, i.e., trust domain. However, the risk never raises from zero, because it is highly unlikely that a private element made open to public.

In addition to investigating the above-mentioned cases, we would like to present the evaluation (figures 1(c) and 1(g)) for an adopter with PI=0.99, who is not absolutely conservative user. We observed the risk of this user is zero. To summarize, except for absolutely conservative adopter, the risk from the proposed C-CBPM model is zero, i.e. learning pricacy policies from community help mitigate privacy risks.

5 Conclusion

The basic premise of this work is that while it may seem counter-intuitive to implicitly trust a friend regarding ones privacy, it is clear that individuals with a good degree of privacy knowledge would be diligent with their privacy settings, and hence encouraging them to share their privacy models with friends who are not as savvy. This can help collectively ramp up a novice users privacy of online information. The model is proved to be performing as expected in real world social networks such as Facebook.

We plan to analyze the feasibility of deploying the C-CBPM framework in cloud-based information systems and further study the implications of trust on user information privacy and security vis--vis service-level interactions.

Acknowledgment. This work was supported, in part, by grants from the NSF (under Grant Nos. CNS- 0619069, EPS-0701890, OISE 0729792, CNS-0855248, EPS-0918970, and IIS-1110868) and Office of Naval Research (Award number: N000141010091).

References

1. Venkata Swamy, M., Ramaswamy, S., Agarwal, N.: Cbpm: Context based privacy model. In: IEEE Social Computing (SocialCom), pp. 1050–1055 (2010)
2. Surowiecki, J.: The Wisdom of Crowds. Anchor Books (2005)
3. Venkata Swamy, M., Agarwal, N., Ramaswamy, S.: Collective context based privacy model. Journal of Ambient Intelligence and Humanized Computing (2012) (to appear)
4. Venkata Swamy, M., Agarwal, N., Ramaswamy, S.: Enhancing privacy using community driven recommendations: An investigation with facebook data. In: Proceedings of the 19th AMCIS Conference, AMCIS 2013. AIS (2013)
5. Einwiller, S., Geissler, U., Will, M.: Engendering trust in internet businesses using elements of corporate branding. In: Proceedings of the 16th AMCIS Conference, AMCIS 2000. AIS (2000)