Poster: Adaptive Disclosure Control System Using Detection of Sensitive Information in SNSs

Shimon Machida [1] Tomoko Kajiyama [2] Shigeru Shimada [3] Isao Echizen [1,4]
[1] The Graduate University for Advanced Studies 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan
[2] Aoyama Gakuin University 5-10-1 Fuchinobe, Chuo-ku, Sagamihara-shi Kanagawa, 252-5258 Japan
[3] Tokyo Metropolitan University AIIT 1-10-40 Higashi Ohi, Shinagawa-ku, Tokyo, 140-0011 Japan
[4] National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan
{shmachid, iechizen}@nii.ac.jp, tomo@ise.aoyama.ac.jp, shimada-shigeru@aiit.ac.jp

1. INTRODUCTION

People of all ages have become regular users of social networking services (SNSs). However, SNS users often unintentionally post a message on SNS, including one's own sensitive information or friends', and is revealed to other users. These unintentional posts can result to unforeseen problems, such as losing a friend or even one's job. Thus, SNS users are sometimes regret after posting a message. In Y. Wang's (2011) research [1], posting a message when in a highly emotional state or inebriated can result in unintentional revelation of sensitive data. In addition to allowing unnecessary access to sensitive information, they have realized with regret that information has been leaked to unintended users [2]. In order to detect potential privacy leaks in any situation such as highly emotional state, and without depending on one's personal opinion, requires making an objective judgment about whether a message includes sensitive information. However, there are no effective guidelines for making such judgments in SNSs. In the following, we described an information classification table for invasion of privacy in SNSs that can be indicator of objective judgment. Moreover, we developed a prototype system that adaptive disclosure control system for Facebook using detection of sensitive information leaks applied the classification table.

2. CLASSIFICATION TABLE

In order to detect potential privacy leaks before posting a message to SNSs, we defined a classification table for invasion of privacy in SNSs [3]. The classification table is based on a guideline for the invasion of privacy in archives of Japan. Archives are an institution that manages and publishes archived documents such as historical documents and official documents. Although its collection in archives are ideally open to the general public as much possible, it holds the potential for including information that has the risk of human rights abuse, such as invasion of privacy, and of unfairly impeding the rights and interests of individuals and corporations. Thus, their guidelines for publishing in official documents cover the invasion of privacy. As one of the attempts, an invasion of privacy information classification table [4] is configured of two axes, with details of information that should be kept private in six categories along the vertical axis and degree of importance of information that should be kept private in three categories along the horizontal axis, together with respective non-disclosure periods. However, the details of the information that should be kept private on the vertical axis include classifications that are unlikely to occur in SNSs, causing a problem in the non-disclosure periods corresponding to the degree of importance of categories on the horizontal axis, in that when a user posts on a popular SNSs, in principle there is real-time access without waiting for the non-disclosure period to expire. Based on this classification table in archives, we define classifications in accordance with the contents of information that should be kept private, reflecting on events where SNS users express regret after posting on popular SNSs [1, 2], and define a version of classification table for invasion of privacy in SNSs in which a non-disclosure period is replaced by disclosure level [3] using Dunbar's circle [5]. Our proposed table is shown in Table 1.

3. PROPOSED SYSTEM

We developed a prototype system that ADCS: adaptive disclosure control system for Facebook applied the classification table. This system focused on the boundary of the disclosure range helps prevent users from unintentionally revealing sensitive information. The user interface is shown in Figure 1.

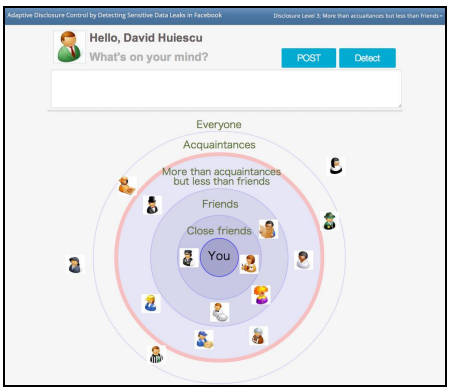

Figure 1. Adaptive Disclosure Control System for Facebook

The post button uses to post a message in the text field to Facebook. The detect button uses to judge about whether the message includes sensitive information before it is posted. The pull down list at the upper right of the window has disclosure levels from 1 ("close friends") to 5 ("everyone"). The use is selection of disclosure range. There are also five circles (Dunbar circles) below the text field. The smallest circle represents the user. The other four circles represent the disclosure levels, from 1

("close friends") to 4 ("acquaintances"). These circles are arranged around the user's circle. The external area beyond the circles corresponds to disclosure level 5 ("everyone"). Pictures of each friend in Facebook are displayed on these circles. Users can easily move the picture of a disclosure target from one circle to another circle using drag and drop. The two main functions are detecting the leakage of sensitive information and suggesting the disclosure range in accordance with a message. To detect sensitive information in a message, the system executes keyword analysis and semantic orientation analysis, which analyze the information using an already created classifier based on a support vector machine. If it contains sensitive information in a message, the system suggests the strictest disclosure level and sends a warning message to the user. As a result, the user can recognize to include sensitive information before it is posted, and revise the message or suggested disclosure level and thereby adjust the disclosure target.

4. CONCLUSION

In this paper, in order to prevent privacy leaks to SNSs, we described the classification table for invasion of privacy in SNSs. Moreover, we developed a prototype system on Facebook that adaptive disclosure control system using detection of sensitive information. We are currently evaluating the table and the system.

5. REFERENCES

[1] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. Leon, and L. Cranor, 2011. I regretted the minute I pressed share: A qualitative study of regrets on Facebook. In Proceedings of the Seventh Symposium on Usable Privacy and Security SOUPS '11. ACM Press

[2] M. Sleeper, J. Cranshaw, P. Kelly, A. Acquisti, 2013. I read my Twitter the next morning and was astonished. A conversational perspective on Twitter regrets. In Proceedings of the 2013 ACM annual conference on Human factors in computing systems CHI '13. ACM Press, pp. 3277-3286

[3] S. Machida, S. Shimada, and I. Echizen, 2013. Settings of Access Control by Detecting Privacy Leaks in SNS. In Proceedings of the Signal-Image Technology & Internet-Based Systems SITIS '13. pp.660-666

[4] Akira Toshima, 2009. Issue of content sharing in Archives. Archives no.35, pp. 40-44

[5] R. IM. Dunbar, 1998. The social brain hypothesis. brain, pp.10

Table 1. Information classification table for invasion of privacy in SNS

Classification and disclosure level according to degree of importance of information that should be kept private / Classification according to contents of information that should be kept private		Disclosure Level 1	Disclosure Level 2	Disclosure Level 3
		Close friends (1-5 people)	Friends (6-15 people)	More than acquaintances but less than friends (16-50 people)
		Communication at least once a week	Communication at least once a month	Communication at least once every six months
		Information that is a particularly important secret about individuals, where there is a danger of unfair impairment of those individuals' rights or profit during their lifetimes if that information were made public	Information that is an important secret about individuals, where there is a danger of unfair impairment of those individuals' rights or profit in social life if that information were made public	Information that is secret about individuals, where there is a danger of unfair impairment of those individuals' rights or profit if that information were made public
Information relating to inner nature of individuals	Beliefs, philosophy	Philosophical opinions of individuals		General social philosophy
	Religion	Religions outlook of individuals		
Information relating to physical and mental states of individuals	Medical history	Illnesses and diseases of individuals (severe)		Illnesses and diseases of individuals (light-to-moderate)
	Physical and mental records		Psychological symptoms of individuals	Physical information of individuals (incl. height, weight)
Basic information and information relating to the life situation of individuals	photographs		Photographs that can identify individuals	
	Domestic situation			Family information such as family structure and domestic situation
	Personal behavior	Aberrant personal behavior, activity status		Day-to-day personal behavior, activity status
Information relating to individuals' background and social activities, etc.	Criminal and unlawful actions	Criminal action (severe)	Criminal action (light-to-moderate)	