A Trust-Aware Tag-Based Privacy Control for eHealth 2.0

Kyle Levy, Brent Sargent, Yan Bai

Institute of Technology, University of Washington Tacoma, Tacoma, WA 98402
levyk@google.com, bsargent@uw.edu, yanb@uw.edu

Abstract

Healthcare social networking is an emerging web 2.0 application that promise to bring about a whole revolution in the way that health care is delivered. In healthcare social networks, health information requires extra protection as its disclosure can have serious repercussions in the content owner's private and professional life. We have developed a prototype of healthcare social network system (called Husky eHealth 2.0) with a trust-aware tag-based privacy control scheme. The scheme protects private information from unauthorized access using both tagging and trust ratings information. A preliminary evaluation of the prototype system is also given. The results have shown that Husky eHealth 2.0 provides an effective privacy control to healthcare social network users.

Categories and Subject Descriptors:
H.0 GENERAL

General Terms: Security

Keywords: eHealth 2.0, Web 2.0, social networks, access control, privacy, security, tags

1. Introduction

Said to revolutionize the way that health care is delivered, eHealth 2.0 is a promising new Web 2.0 technology that allows for an unprecedented level of collaboration between patients, care-givers and physicians. As broadly defined in [4], it is an emerging trend of health-centric blogs, wikis, social networks, tags, tag clouds and podcasts that promise to bring about a whole revolution in the way that health care is delivered. Patients now have the opportunity share treatment experiences with other patients and garner comfort during difficult times. Patients are also able to link up with physicians around the world, who donate their time to helping others within this online community.

Furthermore, physicians are able to share information with other physicians. This level of collaboration in the world of medicine ensures a higher quality of health care for the world, as well as serving as a catalyst to speed up research that has the potential to cure diseases that incapacitate or take the lives of millions around the world.

Recent legislation in the United States allows for incentives for health care providers that exhibit "meaningful use" of certified electronic health record (EHR) systems [3]. It will enable providers to electronically share information about patients. Currently, many states are requiring that healthcare providers be fully converted to EHR systems by 2015. This legislation brings about an even greater potential to expand eHealth 2.0 into other, more intriguing areas. Physicians who donate their time to assisting patients through the eHealth 2.0 channel will now have the ability to view a patient's health records, allowing for more informed diagnoses as well as a much better quality of health care in general.

Amidst this online revolution, privacy control is very crucial for all eHealth 2.0 applications. It deals with highly sensitive health information. However, eHealth 2.0 is the Web 2.0 applied in healthcare, and is at the early stage of development [7]. Only a few major active eHealth websites exist [5]. To the best of our knowledge, many social eHealth websites have not implemented an effective and efficient privacy control.

Our preliminary study shows that tag-based privacy control is a very suitable for addressing the privacy issues present in eHealth 2.0 [1]. Our contribution to this research is developing an eHealth 2.0 website (called Husky eHealth 2.0), implementing a tag-based privacy control for our system and assessing its viability. The rest of the paper is structured as follows: We introduce our Husky eHealth 2.0 system in Section 2 and describe the design of trust-aware tag-based privacy control scheme and the prototype implementation in Section 3, a survey of effectiveness of Husky eHealth 2.0 system and its privacy control scheme is presented in Section 4, and finally conclusion and future work are given in Section 5.

2. Husky eHealth 2.0 System Overview

In this research, a health social networking website has been created. It's entitled "Husky eHealth 2.0", and is shown in Fig. 1.

SIGITE'11, October 20–22, 2011, West Point, New York, USA.
Copyright 2011 ACM 978-1-4503-1017-8/11/10...$10.00.

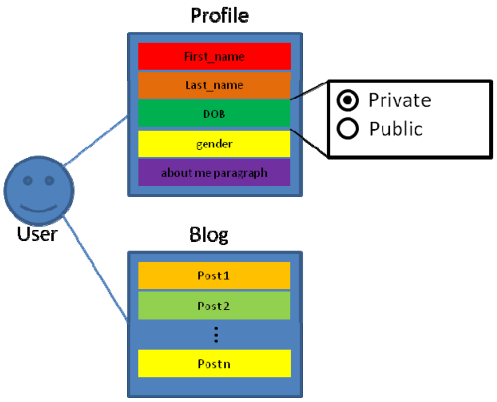


Fig. 1: User and Account Management in Husky eHealth 2.0 system

As shown in Fig. 1, in Husky eHealth 2.0 users can create an account, and once registered, post and edit contents. Each account includes a profile which contains user's first and last name, gender, birthday, and an 'about me' section. In the 'about me' section, user can have an introduction of user themselves. All of this information are optional and can be marked public but are private by default.

Registered users can also create different tags to be applied to themselves in the settings pane. Users can add, delete and modify tags whenever they choose to. In addition, a baseline user trust rating can be specified in the settings pane. This rating indicates how "trustworthy" a user must be in order to view another user's content. Tagging and trust ratings methodologies will be explained in detail in the Section 3.

Registered users are also given a blog in which they can use to create and edit posts. Each post has a title and related content. Tags can be specified for each individual post. Posts can also optionally be made public or private. Registered and unregistered users alike are able to view anyone's profile and blog titles. However, the security settings configured by users determines who can actually see the content. In order to aid easier browsing through the site, an 'All Users' page was created which lists all users registered on the site.

3. Proposed Privacy Control Scheme

The objectives of our privacy control scheme for health social networks are two-fold: 1) protecting the privacy of the social network users, and 2) remaining flexible enough to allow for meaningful interactions between the users. Due to the sensitive nature of the content posted within eHealth social networks, a privacy scheme catering to this type of social network should be more restrictive than other types of social networks, such as Facebook. A user in this context should have the ability to control the privacy of their personal information in a much greater degree of granularity, with content being visible only to users who either have a legitimate need for the information, as well as those who have the potential to provide help in some way. Therefore, the system should also prove to be flexible enough to allow users with the same interests to convene and communicate amongst each other. In a collaborative social network, such as eHealth 2.0, this goal is especially important.

We use two main components: tags and adaptive trust ratings, to achieve privacy protection while maintaining flexibility. In this section, both of these concepts are explained in detail.

3.1 Tags

Tags allow users to categorize both themselves and their content. Fig. 2 provides an illustration of this concept.

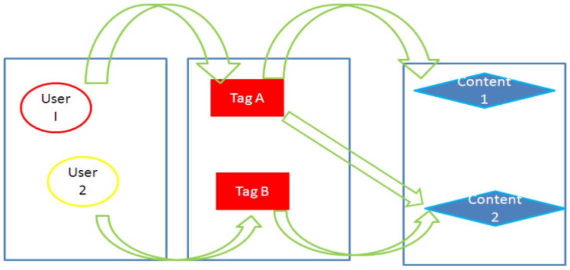


Fig. 2: Tagging in Husky eHealth 2.0

By applying a tag such as 'back pain' to one's user profile, that individual is essentially identifying his or herself as part of a larger group of users who also either have or are in some way interested in back pain. Tags are created when users apply them to their user profiles, and can be called anything they like. This allows users to distinguish themselves from others in a way that is meaningful.

Tags are used as a basic means of restricting information to only relevant parties. In health matters, one might only want to disclose information to those who need it. The concept of tags within the context of the Husky eHealth 2.0 privacy control system allows for this sort of selective disclosure. For example, by applying the tag 'back pain' to a blog post, it displays that blog post only to individuals that have also given themselves the tag "back pain." Likewise, if a user applies the tag 'heartburn' to their user profile, they now have potential access to content also tagged with 'heartburn.' This fact serves to slightly categorize content, but also start to restrict the flow of information across the system to only relevant parties. To allow for additional privacy control, users can also optionally label posts public or private. Neither is required, but if a post is made *public*, it allows unregistered users to view it. If it is labeled *private* it only allows the poster to view.

One of the issues with a tagging system is tag abuse, i.e. users using tags irrelevant to themselves. This can be curbed by a couple methods. First, beyond these basic tags, special tags that require some certification can also be implemented. For example, the tag 'doctor' can be restricted and reserved for only those people who have sent in credentials to be verified. Support groups could have special tags for themselves reserved and restricted to certain members if needed. Another method could take a random survey of users identifying themselves with a large number of tags, and manually check that users aren't trying to just associate themselves with everything in order to attempt to gain access to anything and everything.

3.2 Adaptive Trust Rating Algorithm

Only users that the content owner considers to be trustworthy should be able to view content deemed sensitive by the content owner. The notion of user trustworthiness is an effective way to protect information confidentiality in social networking environment. In this research, we proposed an adaptive trust rating algorithm for privacy control in eHealth 2.0.

The adaptive trust rating algorithm enables content owners to put what could be considered as a selectively-permeable mask on their information of that other users are allowed to see. Only users who the content owner implicitly trusts should be allowed to view their posted content. This algorithm consists of two components: trust metrics and rating calculation. A detailed description of these two components is given in Sections 3.2.1 and 3.2.1.

3.2.1 Trust Metrics

Trust metrics are used to quantify a subjective value, *trust*. Each metric represents a particular value that holds precedence to a user when determining the trustworthiness of another user in a health social networking context; an example of such a trust metric is user availability. The more the time a user logs in an eHealth 2.0 website, the higher the level of trust.

These trust metrics were designed to be modular, in that they can simply be plugged-in as one of the factors used to calculate a user's trust rating. It allows a user some degree of freedom to customize their privacy requirements, so that they accurately reflect their view of trust. A user may also consider one value to be of more importance in determining who they consider to be a trustworthy user. The proposed algorithm enables this freedom for user by allowing them to select from a pool of pre-defined metrics. To allow for further customization, users can define a weight, which comprises a percentage of the maximum possible rating. If no weights are defined by the user, the system compensates by distributing the weight of each metric evenly.

3.2.2 The Rating Calculation

Prior to introducing our trust rating calculation method, we define terminology we used in the method.

- *User Metrics*: We define four metrics, user availability, user popularity, user participation, and user's level of competency to assess a user's performance. User availability is the number of successful login attempts a user has made; User popularity is the number of profile views a user received; User participation is the number of posts a user creates; User's level of competency refers to a user's credibility of their information source.They all measured in a defined period, such as a week.
- *Registration Group*: It is unfair to compare the behavior of a brand-new user to a more seasoned user. For example, one could expect that a more seasoned user most likely contributes more than the new user, due to the fact that they have had the opportunity to create more social linkages (e.g., friends) within the context of a social network, and thus, more than likely, create more posts. To address this situation, our algorithm compares a user to the subset of a social network community whose users have all registered in the same month as the user to be evaluated. This subset of the community is known as a *registration group*.
- *Metrics Scores*: For each metric specified by the content owner, individual user's score and registration group average are logged. The group average is sampled on a periodic basis and creates a model of group behavior for a given metric. It is used by the rating algorithm as a way to compare how a user is performing in a specific area when compared to their peers. In our implementation, we sample the group average weekly. A metric score is defined as the ratio of individual user's score and registration group average.

Assuming a user's (u) trust rating in terms of a given content owner (o), is $T_o\langle u\rangle$, $T_o\langle u\rangle$ is determined by equation (1):

$$T_o\langle u\rangle = C_{MAX}\,((s_1 * w_1) + (s_2 * w_2) + \ldots + (s_n * w_n)) \qquad (1)$$

where s_n is the metric score of u for metric n,, w_n represents the weight of metric n and $C_{MAX} > 0$ is a pre-defined constant representing the maximum trust rating.

Trust ratings are dynamic in nature, allowing content owners to customize them to their liking. They must be calculated upon each content request, which typically occurs when a user loads a page that belongs to another content owner's profile, such as that content owner's blog.

3.2.3 Trust Ratings: A Practical Example

We illustrate the practical use of the adaptive user trust rating algorithm to determine the eligibility of a user to view another user's content. In this example, user Alice has navigated to user Bob's blog, containing several posts that Bob has created. Bob's trust file consists of two trust metrics: user availability, which Bob has assigned a weight of 40%, and user popularity taking up the remaining 60%.

Alice belongs a group of users who have registered in May 2011; this is Alice's registration group, which Alice will be compared against in order to determine her trustworthiness in the eyes of Bob. Let's assume that Alice has achieved the following values in the two areas Bob sees as important:

User Availability (m_1): 4 successful authentication attempts within the past week

User Popularity (m_2): 20 profile views within the past week

Keeping the above information in mind, let us also assume that Alice's registration group has generated the following values for the two above metrics:

User Availability (m_1): 3 successful authentication attempts last week

User Popularity (m_2): 25 profile views last week

Having obtained this information, each of the two metric modules now calculates Alice's "score" in each of the two areas, the score being defined as the ratio of Alice's quantity for one metric to the average quantity achieved by an average individual in Alice's registration group. Thus, Alice's resulting score in each of the respective area would be calculated as follows:

User Availability Score (s_1) = m_1 / M_1 = 4 / 3 ~ **1.33**

User Popularity Score (s_2) = m_2 / M_2 = 20 / 25 ~ **0.8**

With the above information having been calculated by each of the two metric modules, it can now be plugged-in to the rating algorithm to be used in the computation of Alice's trust rating within the context of Bob's content. Working under the assumption that Bob has specified a trust rating requirement of 80 out of a maximum trust rating of 100 and that Alice has a matching tag, Alice's eligibility to view Bob's posts would be determined as follows:

Trust Rating (T_{Bob}<Alice>) = C_{MAX} ((s_1 * w_1) + (s_2 * w_2))
= 100((1.33 * 0.4) + (0.8 * 0.6)))
Max metric value exceeded → = 100(0.532 + 0.48)

Default to max metric value → = 100(0.4 + 0.48)
= 100 * 0.88
= **88**

Notice that in the above case, Alice has exceeded the maximum value for the metric, user availability. This would violate the integrity of the process, the value must be capped at the maximum permissible value for that particular metric. For user availability, this value is 0.4.

Alice has achieved a trust rating of 88, which exceeds Bob's requirement. Bob thus implicitly trusts Alice. A high-level view of the algorithm is shown in Fig. 3.

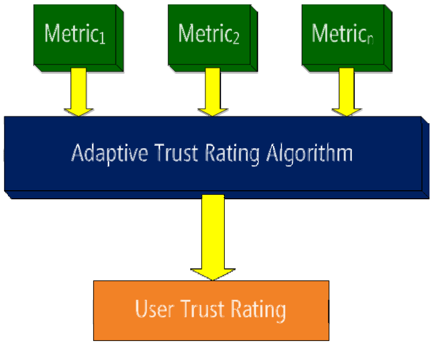
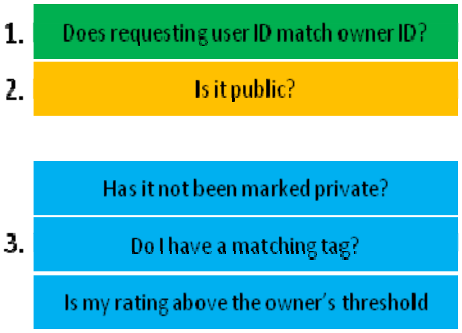


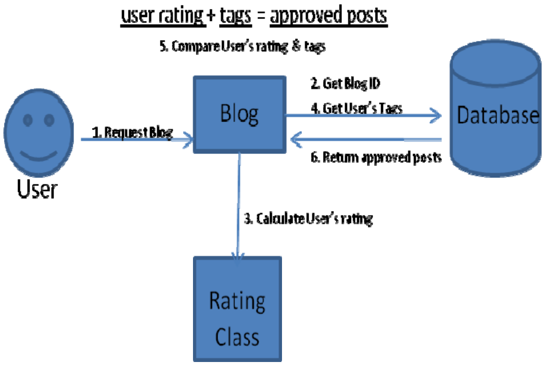
Fig. 3: The Adaptive Trust Rating Algorithm

3.4 User's Content Visibility Testing

In this section, the entire process for determining user content visibility is described. For this example, we assume that the user requests to view a content owner's blog. Fig. 4 illustrates the criteria and steps for user's content visibility testing.


1.
Does requesting user ID match owner ID?
2.
Is it public?
Has it not been marked private?
3.
Do I have a matching tag?
Is my rating above the owner's threshold

(a) Testing Procedure



(b)Testing Flow

Fig. 4 User's Content Visibility Testing

As shown in Fig. 4(a), there are 3 tests to determine if a blog post will be displayed, and passing one test will bypass remaining tests. The first test checks if the user requesting the blog is actually the owner of the blog. If not, then the second test is performed to see if the post is public. If both tests are failed, the third test is conducted. It has 3 criteria: the post has not been marked private, the tag is matched to a tag associated with the post, and requesting user's rating is greater than or equal to the blog owner's threshold. If all three criteria are met, then the post is displayed. If all 3 tests fail, the post is hidden as if it doesn't even exist.

4. SURVEY RESULTS

In our Husky eHealth 2.0 prototype system, the adaptive trust rating algorithm is static in design, with users' trust ratings being derived from the aggregation of all predefined trust metrics, user availability and user participation. It is important to note that the two metrics do not in any way represent the entire set of metrics available to any given user. Rather, these are being used as example metrics to help illustrate the proposed privacy scheme.

In order to determine the correct weight for the two metrics and their viability, we conducted an online survey. In the survey, we also asked users' opinions on other two trust metrics: user popularity and user's level of competency. This helps us to evaluate possible new trust metrics.

In total, 38 individuals responded to the survey, 34 of which indicated that they were current active members of a social network as indicated their login frequency and posting frequency [Fig. 5 (a) and (b)]. In particular, 56% respondents indicates that they login to a social network at least once every day, and 41% respondents indicate that they log in substantially more compared to when they first registered. Moreover, about 90% respondents think user availability is more important than user participation in determining trustworthiness of a user. We thus developed the baseline (e.g., default value and weight for user availability and user participation) according to the survey results.

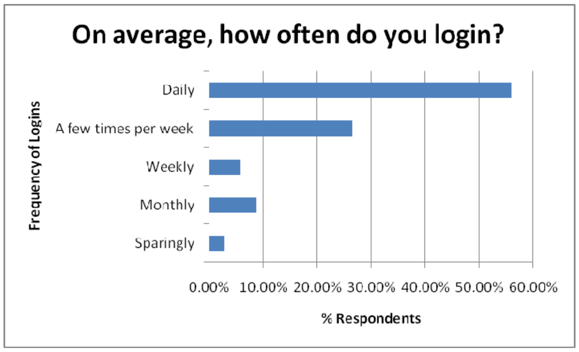


Fig. 5 (a): Average User Login Frequency

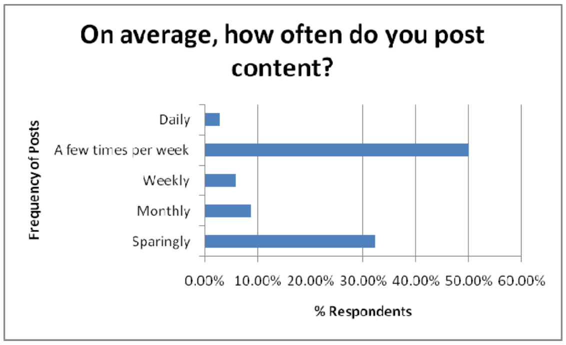


Fig. 5 (b): Average Users' Posting Frequency

Respondent opinions on user competency and user popularity are shown in Fig. 6. It shows that user popularity is more vital than user competency when determining the trustworthiness of a user.

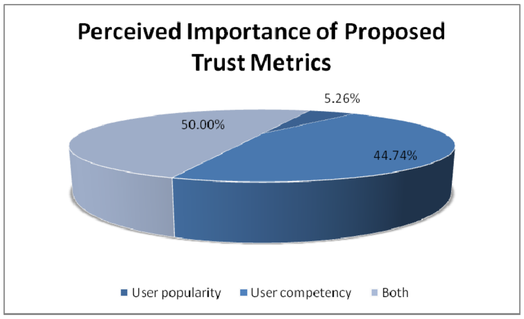


Fig. 6: User Competency vs. User Popularity

The survey also shows that nearly 90% respondents have and will seek online health care information, and place some level of trust in the medical advice provided by online resources [Fig. 7]. The finding indicates that eHealth 2.0 is becoming an important and useful resource for improving healthcare services.

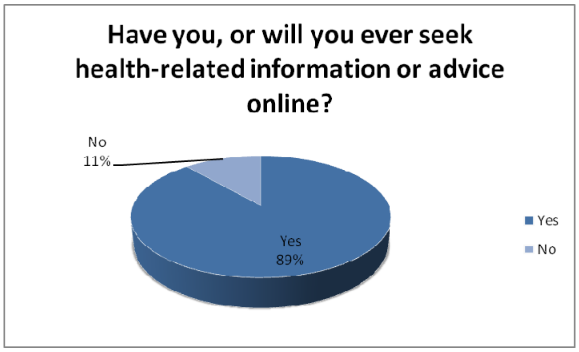


Fig. 7: Usage of Online Health Information Resources

5. CONCLUSION

Healthcare social networking will significantly improve health care services. Privacy control is vital to many kinds of social networking applications, in particular to healthcare social networks. In this research, we have developed a prototype of healthcare social network system (called Husky eHealth 2.0). To enhance our Husky eHealth 2.0 system's privacy control scheme, we considered several important factors in related to the privacy requirements in eHealth 2.0 applications, including user availability, user popularity, user participation, and user's level of competency. We developed and implemented a trust-aware tag-based privacy control scheme based on these factors. We demonstrated via online survey that the Husky eHealth 2.0 provides an effective privacy control to healthcare social network users.

REFERENCES

[1] Y. Bai, X. Su and D. Su, "A Study of Privacy Control in eHealth 2.0", *2011 4th IEEE International Conference on Computer Science and Information Technology (ICCSIT 2011)*, Chengdu, China, June 2011.

[2] M. Hart et al, "Usable Privacy Controls for Blogs," in the *Proceedings of the 2009 International Conference on Computational Science and Engineering*, pp. 401-408.

[3] "Medicare and Medicaid EHR Incentive Program Final Rule", *Center for Medicare and Medicaid Services*, http://www.cms.gov/EHRIncentivePrograms, accessed on August 19, 2011.

[4] H. Wittman and L O'Grady, "eHealth in the Era of Web 2.0", University of Toronto, Canada, http://www.ilit.org/transliteracyweb2/files/ehealtand web2.0.pdf, accessed on August 19, 2011.

[5] J. Hicks and X. Zhong, "Hyoumanity: Social Seach and Incentive Alignment in Health Care", School of Information, University of California - Berkeley, USA, http://www.ischool.berkeley.edu/programs/masters/projects/2009/hyoumanity, accessed on August 19, 2011.

[6] J. Anderson, J. Bonneau, C. Diaz and F. Stajano, "Privacy-Enabling Social Networking Over Untrusted Networks", in the *Proceedings of the 2nd ACM workshop on Online social networks*, pp. 1-6.

[7] Benjamin Hughes, Indra Joshi, and Jonathan Wareham, "Health 2.0 and Medicine 2.0: Tensions and Controversies in the Field", *Journal of Medical Internet Research*, March, 2008, pp. 23-34.