Towards Robust Trust Establishment in Web-Based Social Networks with SocialTrust

James Caverlee
Dep't of Computer Science
Texas A&M University
College Station, TX 77843
caverlee@cs.tamu.edu

Ling Liu
College of Computing
Georgia Tech
Atlanta, GA 30332
lingliu@cc.gatech.edu

Steve Webb
College of Computing
Georgia Tech
Atlanta, GA 30332
webb@cc.gatech.edu

ABSTRACT

We propose the SocialTrust framework for tamper-resilient trust establishment in online social networks. Two of the salient features of SocialTrust are its dynamic revision of trust by (i) distinguishing relationship quality from trust; and (ii) incorporating a personalized feedback mechanism for adapting as the social network evolves.

Categories and Subject Descriptors: H.3.5 Information Storage and Retrieval: Online Information Services

General Terms: Algorithms, Experimentation

1. INTRODUCTION

Web-based social networking services like the ones offered by MySpace and Facebook support the management of social relationships, connecting millions of users. MySpace alone has grown from 1 million user accounts in 2004 to an astonishing 250 million accounts today.

This growth has not come without a price, however. The large social networking sites have been the target of specialized phishing attacks, impersonating profiles, spam, targeted malware dissemination, and new threats are certain to emerge as attackers grow in sophistication. While there are important problems associated with securing the social network infrastructure, we explore vulnerabilities to the quality of information available through online social networks even when the underlying social network infrastructure has been secured. In particular, we identify three vulnerabilities:

- **Malicious Infiltration:** Most online social networks provide some limits as to who can participate, often requiring a valid email address or a registration form. As a result, many social networks give the illusion of security [1], but malicious participants can still gain access.
- **Nearby Threats:** The small world phenomenon [6] means that there is a short distance in the network between any two participants. Even if a user has tight control over his direct friends, malicious users can be just a few hops away.
- **Limited Network View:** Even if a user in the social network maintains tight control over her friends *and* closely monitors the quality of her neighbors' friends, she will still have access to only a limited view of the entire social network, meaning users have no assurances over the vast majority of all participants in the network.

Malicious users can exploit the perceived social connection between users for increasing the probability of disseminating misinformation, of driving participants to the seedy side of the Internet (e.g., to sites hosting malware), and of other disruptions to the quality of community-based knowledge.

2. THE SOCIALTRUST MODEL

With these problems in mind, we present the initial design of SocialTrust, a reputation-based trust aggregation framework for supporting tamper-resilient trust establishment in online social networks. The benefits of reputation-based trust from a user's perspective include the ability to rate neighbors, a mechanism to reach out to the rest of the community, and some assurances on unknown users.

Initially all users are treated equally. SocialTrust supports trust maintenance through dynamic revision of trust ratings according to three critical components: the current quality component of trust $Tr_q(i,t)$, the history component, and the adaptation to change component. The SocialTrust score for user i at time t is defined as:

$$ST(i,t) = \alpha \cdot Tr_q(i,t) + \beta \cdot \frac{1}{t}\int_0^t ST(i,x)dx + \gamma \cdot Tr'_q(i,t)$$

where $Tr'_q(i,t)$ is the derivative of $Tr_q(i,x)$ at $x = t$. This approach is similar to a Proportional-Integral-Derivative (PID) controller used in feedback control systems [7].

By tuning α, β, and γ, the SocialTrust model can be optimized along a number of dimensions, e.g., (i) to emphasize the most recent behavior of a user in the network (via higher values of α); (ii) to de-emphasize the current user's behavior in the context of his entire history of behavior (via higher values of β); or (iii) to amplify sudden fluctuations in behavior (via higher values of γ).

Given the overall SocialTrust approach, what is an appropriate choice of the base trust metric $Tr_q(i,t)$? In light of the vulnerabilities previously identified, we suggest that a good base trust metric should incorporate two key features: *1. Distinguishing Relationship Quality from Trust.* Many trust models (e.g., [4, 5]) evaluate the relative trustworthiness of a node (or user, in our case) based on the trustworthiness of all nodes pointing to it, but make no distinction about the relationship (or link) quality of each node. In essence, these approaches make no distinction between the trust placed in a user and the trust placed in a user's relationships. Intuitively, we would like to differentiate between

WWW 2008, April 21–25, 2008, Beijing, China.
ACM 978-1-60558-085-2/08/04.

users who consistently engage in high-quality relationships with other users versus users who tend to engage in lower quality relationships.

2. Incorporating Personalized User Feedback. Second, trust models based solely on network topology are divorced from the underlying behavior of the users in the network. Relationships in the online social network provide the basis for trust aggregation, but there is no feedback mechanism for dynamically updating the quality of the trust assessments based on how well each user in the network behaves. Hence, we are interested in "closing the loop" so that the trust assessments may be dynamically updated as the social network evolves and as the quality of each user (with respect to user feedback) changes over time.

Based on these observations, SocialTrust assesses user i's trust rating $Tr_q(i)$ according to the user's relationship link quality $L(i)$ and her feedback rating $F(i)$ through a recursive formulation:

$$Tr_q(i) = \lambda \sum_{j \in rel(i)} L(j) \cdot Tr_q(j)/|rel(j)| + (1-\lambda)F(i)$$

where $|rel(i)|$ is the total number of relationships i participates in and λ is a tunable mixing parameter. The intuition is that a user's trustworthiness should be determined by both: (i) the number and trustworthiness of the users who recommend her (via relationships in the social network); and (ii) the relationship link quality of each recommending user. In this way, a recommendation from a high-trust/high-link-quality user counts more than a recommendation from a high-trust/low-link-quality user. The feedback rating $F(i)$ favors users who have been rated highly by other users, according to the mixing factor $1 - \lambda$.

3. PRELIMINARY EVALUATION

We have evaluated SocialTrust in a simulation over a directed graph consisting of 5,199,886 nodes and 19,145,842 relationship links that we harvested from MySpace. We refer the interested reader to [3] for more discussion of the simulation setup and how link quality and feedback are computed.

We consider a scenario in which an *originating user* has an information need (e.g., looking for a job in Texas, finding a good restaurant) for which she can use her social network. The basic scenario is this: a user browses her relationships up to some radius looking for candidate users to ask; based on an analysis of their profiles, she constructs a set of candidate users who might satisfy her information need; based on the provided trust ratings, she selects the top-k most trusted candidate users; she asks all top-k; if she is satisfied, she provides positive feedback to the trust manager; otherwise, she provides negative feedback. We model two types of users: (i) malicious users, who always provide an irrelevant response when asked; and (ii) legitimate users, who sometimes accidentally provide an irrelevant response when asked. For a query q, let R^+ denote the set of relevant users for q throughout the entire space of users and let R_n denote the n top-ranked candidate users (by trust value). We measure a focused version of the standard precision measure that considers the quality of the responses in the top-n (the relative precision @ n): $prec_n = \frac{|R^+ \cap R_n|}{min(|R_n|,n)}$.

In Figure 1, we compare SocialTrust to several related trust models adapted from the Web and P2P domain to online social networks. We consider a PageRank-based trust model that considers only the relationship structure of the social network; a TrustRank-based model that uses feedback ratings as a priori trust (which is equivalent to EigenTrust from the P2P domain); a preliminary SocialTrust [Cred Only] model that incorporates relationship link quality only but no feedback ratings (which is similar in spirit to credibility-based link analysis explored in the Web domain [2]); and the final SocialTrust model. When a proportion of highly-trusted users behave maliciously, PageRank and TrustRank have no mechanism for correcting this bad behavior. In contrast, the SocialTrust model incorporates link quality and feedback ratings into the trust assessment so that bad behavior is punished, and so the resulting precision measures are resilient to the presence of a large fraction of malicious users in the network. These initial results are encouraging, and we working to further explore the key properties impacting SocialTrust.

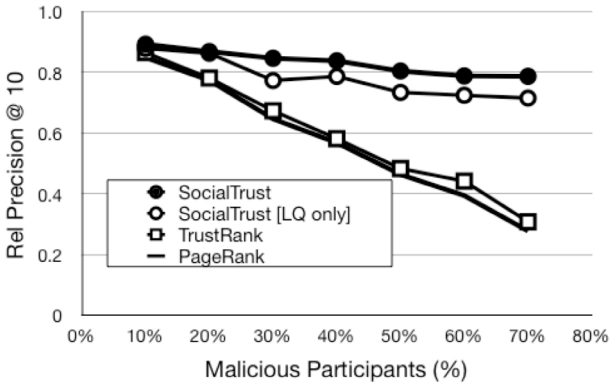

Figure 1: Comparing trust models

4. CONTINUING WORK

In our future work, we are interested in developing context-aware extensions of SocialTrust so that the network may support multiple trust views of each user depending on the context. We also see opportunities to augment the evaluation of relationship link quality, so that it considers more sophisticated features like the nature, duration, and value of each relationship. On the implementation side, we continue work on a SocialTrust-powered community platform that can be layered on top of existing social networks.

5. REFERENCES

[1] S. B. Barnes. A privacy paradox: Social networking in the United States. *First Monday*, 11(9), September 2006.
[2] J. Caverlee and L. Liu. Countering Web spam with credibility-based link analysis. In *PODC*, 2007.
[3] J. Caverlee, L. Liu, and S. Webb. SocialTrust: Tamper-resilient trust establishment in online communities. Technical report, Texas A&M University, 2008.
[4] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating Web spam with TrustRank. In *VLDB*, 2004.
[5] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The EigenTrust algorithm for reputation management in P2P networks. In *WWW*, 2003.
[6] S. Milgram. The small-world problem. *Psychology Today*, pages 60 – 67, May 1967.
[7] M. Srivatsa, L. Xiong, and L. Liu. TrustGuard: Countering vulnerabilities in reputation management for decentralized overlay networks. In *WWW*, 2005.