A Privacy-Aware Architecture for Sharing Web Histories*

Alex Tsow
Shreyas Kamath
L. Jean Camp
{atsow, sskamath}@indiana.edu, ljean@ljean.com

Abstract

Net Trust exposes fraudulent web sites by combining individual opinions and browsing histories over self-selected social networks. This paper describes the security and privacy design choices for the first public-use implementation of the *Net Trust* web rating system. Net Trust differs from prior rating and recommendation systems which leverage peer production and social networks because it is engineered to withstand Sybil attacks, corruption by producing bogus ratings en masse. The system simultaneously strives to ensure privacy with *linking resistance*, *social network confidentiality*, and *account deniability*. Our implementation strikes a compromise between data availability and structurally imposed privacy through a rich-client/lightweight-server architecture. This paper analyzes Net Trust's participants, attackers, security and privacy goals, and implementation choices.

1 Introduction

Net Trust is a rating system that combines the browsing habits and opinions of friends to identify online fraud, phishing and sleaze. Individual web histories and social networks are sensitive data for many. Much media attention has been shed on people who have been fired for online blog postings, pictures, and other personal content. Perhaps more insidious is the back-door trade of private information via personalization services [4] for consumer profiling [22]. Unlike many social referral systems, browser toolbars, web portals, and other free Internet services, Net Trust does not extract payment from its users through analysis, profiling, or reselling of their private data. For instance, the *Alexa* toolbar — which ranks sites, detects fraud, and makes related link recommendations — transmits and logs web histories, search terms, product purchases and other personally identifiable information [1]; Alexa profits from selling aggregate reports of this data, while its corporate parent, *Amazon.com*, attaches portions of collected data to personally identifiable information. We have designed Net Trust to protect its users from attacks on their sensitive information and ratings integrity by malicious peers, dishonest web sites, and even the Net Trust servers.

Net Trust's security and privacy properties are intended to stop the for-profit compromise of ratings and private information. Economic gain motivates malware's ubiquity and sophistication far more than bragging-rights and malicious character-destruction. The strength of its security and privacy properties is tuned to eliminate systematic abuse by profit-seeking agents. In particular, this falls well short of *computational intractability* as required by cryptographic operations; their security and privacy protections rest on financial calculations rather than algorithmic feasibility.

The remainder of the paper is organized as follows: Section 3 presents the Net Trust rating system and illustrates its use against two fraud scenarios. Section 4 defines Net Trust's potential attackers and the system's security and privacy goals. Section 5 presents the system's implementation. Section 6 discusses how well this architecture achieves the security and privacy goals. Section 7 concludes the paper with a summary of properties achieved and directions for improvement.

*This work has been funded, in part, by Google and The Institute for Information Infrastructure Protection (The I3P).

3.2 Social networks

Although a level of protection arises from analyzing an individual's browsing habits, utility improves when individual ratings are combined with ratings from friends. Unlike many social referral systems which infer social networks from centralized analysis of individual behavior [19, 24, 1], Net Trust users form their social network by explicit invitation. Only immediate friends have a direct impact Net Trust ratings — although friends of friends have an indirect influence by changing the behavior of immediate friends.

Net Trust reports the following aggregate data over the direct friends in the toolbar: average of negative ratings, average of positive ratings, number of comments, and number of ratings. User interface testing has shown that this information, when suitably displayed, changes trust behavior [13]. Full ratings — the comments and numerical rating for each friend in the network — are available upon request in a separate window.

In general, a Net Trust user should maintain several different ratings profiles, or *personas*. Each persona has a different set of friends. This is important from both a privacy point of view and a utility point of view. One should not mix web ratings between a persona representing one's professional life and a persona representing one's recreational activities. Moreover, these interests are likely to be properly informed by different friends who visit different web sites.

3.3 Third-party sources

Sometimes an individual's social network is not rich enough to effectively inform trust behavior. *Third-party sources* can help to seed opinion. Third-parties can give one of three ratings — negative, neutral, or positive — to a web site. This rating is indexed entirely by the web site's domain name. The ratings do not mix with the aggregate report from the social network. As with the end-user's social network, individuals get to choose which third-party rating sources they wish to include. For now, the space of third-party sources is limited to those chosen by the Net Trust developers, including the FDIC, BBB, and Site Advisor.

4 Security and Privacy Goals

Net Trust requires several kinds of sensitive information to produce its ratings. Among these are partial web histories, personal identifiers, buddy identifiers, and authentication credentials. Without appropriate protections, this information is subject to abuse. This section identifies the Net Trust participants, their roles, and the system's security and privacy objectives.

4.1 The Participants

There are four principal agents in the Net Trust system: *rating-subjects*, *peer-producers*, *third-party sources*, and *rating-servers*.

The **rating-subject** is the web site which Net Trust evaluates. Their incentives are to promote their own ratings and possibly demote their competitor's ratings. Rating-subjects can host arbitrary content (esp. scripts) — a potential automated-attack vector. Since implicit peer-ratings vary according to visiting patterns (Section 3.1), a web-site may try to automatically influence this with scripted content, e.g. scripted redirects, reloads, and pop-behind windows. Rating-subjects also have control over the structure of their web-site and may manipulate it to confuse the binding of ratings to pages. Moreover, while web sites do not explicitly handle ratings in their role as rating-subjects, they may play strategically as peer-producers.

The **peer-producer** is the essential rating unit in Net Trust. As explained in the Section 3.1, peers alter subject ratings implicitly through their visiting patterns and explicitly through comments or manual evaluation. A social network is defined by one *persona* — a peer account controlled by the end-user — and a collection of *buddies* — the peer accounts that combine with the persona account to rate subjects. In the current Net Trust implementation, the client maintains persona ratings and writes this data to rating-servers. In principle, a malicious client could write arbitrary ratings to the rating-server for a given persona.

personas. For instance, a peer-persona that represents the end-user's professional interests may carry history that should be kept separate from a persona which represents the end-user's recreational interests.

Least necessary URI: In general the page rendered by a browser is function of the URI, HTTP post data, server-state, client-state (e.g. cookies, cache, browser history, browser plug-ins) and protocol details (e.g. IP address). The complete download details reveal far too much information. Net Trust's goal is to supply the least detailed URI that identifies fraudulent web sites.

Least necessary time: Net Trust evaluates web sites based on the visit times and places from peer histories. Too much data in the ratings file and too frequent an update to ratings servers could betray timely knowledge of browsing habits to both the ratings servers and the peers. For instance, a malicious peer could script repeated download ratings files to discover when they've changed. If peers update their central ratings upon browser close or open, then this scripted download can show when someone is at their computer. Similarly, if there are central record accesses and updates for each visit (an option in the Google Safe Browsing toolbar [10]), the server and malicious peer could deduce that a target peer is looking at a specific page at a specific time after a short latency.

5 Design

5.1 The Data Aspect

The Net Trust system relies on peer ratings, third-party recommendations, and social connections to produce its evaluations. The delegation of data storage (and more generally access access) over the Net Trust participants is critical to its security and privacy properties.

Anonymous but consistent peer identification: Net Trust uniquely identifies peers with random-looking *reference-keys*. The hash — a one-way mapping of strings into 128-bit values — of a user-asserted nym (a nickname which may not be unique), email address, and random number generates the Net Trust reference-key. These assertions are not subject to verification, however the hashing regimen serves as a *commitment* to the specified information and prevents someone who only knows the reference-key from recovering the nym and email address.

Local social-network storage: Each Net Trust client manages its own social network. A locally stored file contains reference-keys, email addresses, nyms, and shared secrets (for write authentication with ratings-servers) for each of the end-user's peer-personas. Net Trust defines a persona's social network with a list of immediate friends. The list contains the reference-key *and* generating information (the nym, email address, and random number) for each friend. The same buddy may appear in multiple personas. This local file is the only record of social networking data in Net Trust. The system does not share it with ratings servers or other peers. There is no global view of the social network.

Peer-ratings: Net Trust generates peer-ratings locally but shares them via the rating-servers. The top-level contains the peer's reference-key — but not its generators — and a date indicating its last modification. The children of the top-level structure hold truncated URIs which store the web site's domain (minus the leading `www?.`, if present) and the top-level directory, if present. The record contains no CGI variables, individual pages or script-references. Truncated URIs have either a *visit-node* or an *explicit-rating-node* and an optional *comment-node*. The visit-node stores the dates of initial and last visits as well as the visit count from one to five. When an explicit-rating-node is set, it stores only positive or negative value rating from one to five; it stores no dates or visit-count. The comment-node stores text for a one-line comment.

Third-party recommendations: Third-party recommendations sort web-sites into three categories (positive, negative, neutral) according to their domain name only. Third-parties may specify one of three ratings as its default classification. The recommendation file stores a descriptive identifier of the third-party source, a default rating, and a list of `<domain>,<rating>` pairs.

fields into the client's "Add Friend" interface. Net Trust validates that $h(<\texttt{rand}>, <\texttt{email}>, <\texttt{nym}>)$ equals the reference key. This catches entry errors, but also prevents someone from claiming that a reference key belongs to a different user than the one asserted during registration.

The synchronization module is the only part of Net Trust that interacts with rating-servers. It abstracts read and write process, accessing the local filesystem and/or the rating-servers as necessary. When the user changes from one persona to another, Net Trust calls on the synchronization module to save the first persona history. Early versions saved the history locally followed by a write to the rating-server. Current versions schedule server-writes for later to mask timing and frequency of user activity. Similarly, the synchronization manager loads the rating-sets of a specified social network. The local copy of client-controlled rating-sets is considered authoritative, however buddy ratings must be downloaded from the rating-servers to stay current. As with writes, the early version simply downloaded all requested buddy ratings at the time of request. The current version staggers the update of buddy ratings to mask user behavior and social networks.

Net Trust is a web browser toolbar which displays the aggregate peer ratings (average negative ratings, average positive ratings, number of comments) and up to five third-party classifications. An on-demand external window shows full evaluations by enumerating ratings and comments for each buddy in the network.

6 Discussion

6.1 Security Properties

One of the core properties of Net Trust's social-network-based ratings system is its intrinsic *resistance to Sybil attacks*. Malicious peers must first be accepted as buddies in the victim's social network in order to influence ratings. Moreover, there must be sufficient penetration to override honestly-generated ratings. Committing mass ratings-fraud requires controlling lots of peer accounts *and* large-scale infiltration of social networks. Thus the Sybil attack must target victims individually. Contrast this with open-access search-engine manipulation, where rating-up (or down) a page with bogus links produces a global effect.

By itself, the per-victim-cost property deters mass manipulation with bogus peer accounts. However Net Trust further increases the per-victim-cost with its *identity commitment* scheme. When registering an account, the peer must commit to an email address and a nym. If a Sybil attack were to attempt large scale social engineering by self-inviting into victim social networks, the messages must offer socially meaningful nyms and email addresses. For instance, if Jane Doe and Don Jones are friends, an attacker who wants to join Don Jones's network would maximize the chances of success by spoofing Jane Doe's identity. Perhaps the attacker would register a Net Trust account with the email address, `janedoe69@free-email.org`, and nym, `JaneDoe`. This account will help the attacker compromise Jane Doe's friends, but work against compromising others. Since bogus peer accounts must be tailored to their victims, the social-engineering Sybil attack uses large numbers of Net Trust accounts.

The rating-servers, which grant account resources, can throttle this kind of excessive account creation with proof-of-work [18] schema. The problem could be made even more expensive by automated client-side validation of email invites. At the most rudimentary level, validation compares the email address commitment with the sender address. A more sophisticated version adds a header check for valid domain keys [12]. A more automated invitation process could implement a protocol that sends request and acknowledge email messages. This would force the attacker to control a matching email account for each bogus Net Trust peer. Of course, a sophisticated end-user could execute such policies by hand.

Rating-servers form the central distribution system for third-party and peer data. The servers are entrusted to faithfully transfer data among peers; there are no provisions for peer *read integrity*. However, the Net Trust certificate authority issues signing certificates to third parties who want to maintain lists on rating-servers. The Net Trust client ensures that third-party updates use same signing key as the previous version. This scheme implements read integrity by preventing rating-servers from altering the contents of third-party recommendations.

Write authentication is achieved through secret sharing at the time of account registration. The Net Trust client follows the HMAC protocol for authenticated server writes [17]. The client leverages SSL

friends. The naive synchronization protocol, which updates all buddy-ratings at the time of persona load, trivially discloses the network to a server which logs access times — it simply looks for clustered access. The current synchronization mechanism counters this by randomly scheduling updates over a parameterized time window. This number is adjustable to accommodate a changing anonymity set — a value that will grow as Net Trust gains more users. The deferred update policy does not eliminate the attack, but makes it harder to execute. The diligent attacker counters by performing intersection attacks on aggregate reads over largish time frames.

The Net Trust synchronization policy and history format limits the tracking of user activity. The most invasive level of monitoring would record the time of every visit and send high-frequency low-latency updates to rating-servers. Net Trust only records the initial visit date, last visit date and visit count (a value bounded by 5). Server write updates are parameterized by frequency, with a default setting of 1 day.

A third-party participant — the large email providers — could also pose a privacy threat, if clear-text email remains the principal medium for invitation. Large email providers can infer Net Trust friends and identities by reading the information directly out of an invitation. In large part, they already know a substantial portion of the social network due to header tracking. This privacy loss is a component in the targeted advertising which supports the services. However, attaching browsing habits to this data is beyond the comfort level for many users.

The rating-servers make no authentication demands for those who wish to read peer-ratings. Since reference-keys span a randomly distributed 128-bit space, a download of all ratings by exhaustive search is impractical. Third-parties who want to access peer records without permission must obtain lists of valid Net Trust reference-keys. This may be possible for a powerful entity such as a large email provider. Social engineering — e.g., through bogus invitations — could uncover others. For now, Net Trust credentials are vulnerable to snooping since the client does not encrypt files. Malware could extract reference keys, but in most cases independently control the infected machine — a far more serious problem. In addition to malware, unencrypted account credentials become a problem for poorly configured peer-to-peer file-sharing clients [15]. As with malware, the privacy concerns are far graver than disclosure of Net Trust credentials when large portions of one's hard drive become searchable over a file-sharing network.

7 Conclusion/Future Work

We have implemented an operational prototype which shares web histories among self-selected social networks to create independent trust signaling for web sites. Our prototype implements the rating system which resists manipulation through Sybil attacks. *Identity commitment*, *server authentication*, *writer authentication*, *server-to-client encryption*, and *third-party signatures* ensure ratings integrity. While based on individual web browsing habits and opinion, Net Trust ensures privacy with *linking resistance*, *confidential social networks*, *account deniability*, and limited disclosure of browsing details.

Although the prototype system achieves many security and privacy properties, there are several future directions we wish to develop:

- Net Trust can extend its fraud detection to fight *pharming* (DNS spoofing) by including web-server IPs and certificates in the browsing history. This simple extension could become an important audit for local DNS servers.

- Invitations in Net Trust depend on external channels. Although we envision email as the principal invitation medium, it becomes a privacy liability in the presence of email providers who parse message content. Future work will explore peer-to-peer channels for invitations.

- As implemented, rating-servers are trusted to correctly relay peer-ratings. Future releases should include read integrity or auditing mechanisms for this transmission.

- In order to protect social networks, rating-servers do not authenticate users. Thus unauthorized users who know an individual's Net Trust reference-key can download histories. Although the *account*

[17] H. Krawczyk, M. Bellare, and R. Canetti. Hmac: Keyed-hashing for message authentication. RFC 2104, February 1997.

[18] D. Liu and L. J. Camp. Proof of work can work. The 2006 Workshop on the Economics of Information Security *(WEIS 2006)*, 26-28 June 2006.

[19] B. Markines, L. Stoilova, and F. Menczer. Bookmark hierarchies and collaborative recommendation. In *Proceedings of The Twenty-first National Conference on Artificial Intelligence (AAAI-06)*, 16-20 July 2006.

[20] N. McFarlane. *Rapid Application Development with Mozilla.* Prentice Hall PTR, 2003.

[21] Netcraft anti-phishing toolbar. Accessed 15 June 2007, `http://toolbar.netcraft.com/`.

[22] A. Odlyzko. Privacy, economics, and price discrimination on the internet. In *ICEC '03: Proceedings of the 5th international conference on Electronic commerce*, pages 355–366, New York, NY, USA, 2003. ACM Press.

[23] M. Security. Phishing filter: Help protect yourself from online scams. Accessed 15 June 2007, `http://toolbar.netcraft.com/http://www.microsoft.com/protect/products/yourself/phishingfilter.mspx`, 2007.

[24] Stumbleupon ≫ privacy policy. Accessed 15 June 2007, `http://www.stumbleupon.com/privacy.html`, 9 May 2007.

[25] Y. Zhang, S. Egelman, L. Cranor, and J. Hong. Phinding phish: Evaluating anti-phishing tools. In *Proceedings of the 14th Annual Network & Distributed System Security Symposium (NDSS 2007)*, San Diego, CA, March 2007.