Reliable, Usable Signaling to Defeat Masquerade Attacks

L Jean Camp
Associate Professor
School of Informatics
Indiana University
Bloomington, IN

March 23, 2006

1 Introduction

In the nineties the disconnection between physical experience and the digital networked experience was celebrated - individuals were said to move into cyberspace, become virtual and leave the constraints of the physical realm. Despite the very real existence of tracking and surveillance software, it remains the case that identity assertions online remain problematic at best. While there are many benefits to this relative anonymity online, it also creates a serious problem of distinguishing between valid merchants and criminal enterprises, between reliable web sites and sites that install malware. The paucity of information exacerbates poor understanding of the risks involved in particular transactions, exposing users to losses and driving out parties who would otherwise engage in productive behavior.

This paper describes a mechanism to create highly usable economic signals that enable users to evaluate sites on the Internet, and in particular to specifically identify masquerade attacks. By integrating both peer production and centralized information, the system utilizes both personal local histories and centralized information sources. Limiting the distribution of personal histories to user-defined social networks enables users to constrain and control their own information.

2 Theory & Motivation

Net Trust integrates third party information and data from social networks to create signals integrated with browsing. Signals are information that is difficult to falsify and thus can be used to distinguish between types of otherwise indistinguishable goods. In this case the "goods" in question are web sites. Net Trust communicates structural information from social networks to create difficult-to-falsify signals. These signals will indicate that a web site has a history of reliable behavior, just as good grades indicate that a potential employee has a history of hard work.

As example of an attack that is enabled by lack of signals is a phishing attack. Phishing is difficult to prevent because it preys directly on the absence of resource identification information in trust decisions online. Absent any information other than an email from a self-proclaimed bank, the user must to decide whether to trust a website that looks very nearly identical to the site he or she has used without much consideration. Simultaneously, there is very little that an institution can do to show that it is not a masquerade site.

A second set of attacks that could be decreased with this system are web sites that download malicious code, or exploit browser vulnerabilities to create zombies. For example, a study by Microsoft using monkey spider browsers (browsers which spider the web but act like humans) found 752 sites that subverted machines via browser vulnerabilities. [30] Net Trust is designed to easily integrate such a list, and inform users of the risks of the site, perhaps even interrupting the connection with a carefully designed warning.

At least in part, Internet fraud is enabled by a lack of reliable, trusted sources of information. Such fraud is a large and growing problem. [7] [28] The Federal Trade Commission has reported that in 2004, 53% of all fraud complaints were Internet-related with identity theft topping the list with 246,570 complaints, up 15% from last year. [8] PEW has noted that 68% of Internet users surveyed were concerned about criminals obtaining their credit card information, while 84% were worried about compromise of other personal data. [28]

Although fraud and misinformation exist off the Web, these threats can be somewhat mitigated offline through mechanisms that utilize physical presence, identities, well-defined roles and signaling in physical communities. In the physical realm customers can use visual, geographical and tactile cues that indicate a merchant's professionalism, competence, and even trustworthiness. [22][20] In e-commerce, parties to a transaction commonly are geographically, temporally, and socially separated. [15][16]

Consider the two comparisons in Figure 1. These are both places where one might purchase pearls. Were these markets meters, as opposed to continents, apart there would still be no way to confuse the two. In economic terms Tiffany's has the higher quality and is able to signal this quality through the construction of an impressive facade, location at a prestigious address, and a highly ordered self-presentation. In contract, the signaling in the Ladies' Market indicates high competition, low overhead, and strong downward pressure on prices. In the Hong Kong market, merchants may assure buyers that the pearls are real, perhaps even harvested in Japan. The buyer may be assured that the low prices are

Figure 1: Compare the Entry to the Ladies' Jewelry Market, HK to that of Tiffany's, NY

a result of once in a lifetime opportunity, and that the buyer should pay a premium for this rare chance at owning such high quality pearls. The overall context of the transaction provides information useful in evaluating these claims.

Online these virtual sites would be distinguished only by the web site design, domain name, and corresponding SSL certificates. Imagine one of the merchants in the Hong Kong were named Tifanny. In February 2006, Tifanny.net is available for tens of dollars. Even with a PKI linking the domain name to the merchant's name using an SSL certificate, confusion would be possible. In contrast, brick and mortar businesses can invest in physical infrastructure and trusted physical addresses to send signals about their level of prestige, customer service and reliability. For example, a business on the fiftieth block of Fifth Avenue (arguably the most expensive real estate in New York and thus America) has invested more in its location than a business in the local mall that has in turn invested more than a roadside stall. The increased investment provides an indicator of past success and potential loss in the case of criminal action. Such investment information is not present on the Internet. The domain "tifany.us" is currently available, but creating an equally believable offline version of Tiffany's requires far more investment.

To emphasize the point, consider Figure 2. One is Sun Trust Bank. The other is a computer in Columbia University that was controlled at the moment of that screen shot by a criminal entity (quite possibly on another continent). There is no mechanism for the bank to signal to the virtual customer its investment and thus its quality and authenticity. Any signaling is limited to the form of mass-produced easily copied images (e.g., TRUSTe or BBB trust seals) or hard to understand SSL certificates.

Ironically for an information infrastruture, it is easier to tell apart two jewelry stores offline than it is to distinguish between a bank and a criminal hide-out online.

Security experts immediately recognize the false site. Many users also recognize the false site, based on its domain name and the lack of the icon indicating a SSL certificate. These cues can be falsified, as with registration of confusing domain names (as is the case here

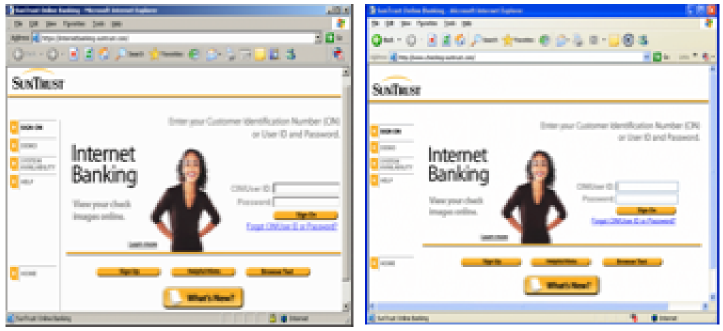

Figure 2: Compare the Portal to Sun Trust Bank to that of an Organized Crime Syndicate

with checking-suntrust.com) and SSL-enabled phishing. Only the SSL icon indicates that the individual is in a low trust environment, because in this case the phisher has purchased a domain name that does not provide distinguishing information. The easy falsification is possible because there are no signals.

The goal of Net Trust is to allow individuals to make better decisions about online resources, by embedding signals into browsing. Net Trust makes these signals unique to each user. Net Trust is more secure than easy to copy visual cues (e.g., the TRUSTe COPA seal or the Better Business Bureau seal) and more understandable than certificates. Net Trust is expandable, and is to be distributed with a BSD license.

2.1 Motivation

Economics of information security is a fairly recent area of formal research. Much of the economics of information security research has been focused on the analysis of extant mechanisms and the modeling of exchanges of vulnerabilities. Even proof of work was designed from an economic idea not from a microeconomic analysis. [11] Indeed, when proof of work was subject to economic analysis the assumptions about producer costs proved to be fatally flawed, as the existence of zombies results in very different production frontiers for spammers and producers of legitimate email. [18] This is the first constructed anti-phishing protocol that has as its grounding a microeconomic model of the system, and unifies technical design, microeconomic models and usability testing.

In 2000 the Computer Emergency Response Team at Carnegie Mellon proposed the Hierarchical Holographic Model. This was the first multifaceted evaluation tool to guide security investments based on economics and risk science. [19] Currently CERT has completed a suite of mechanisms for risk assessments. This systematic approach can be appropriate, depending on the size and expertise of the organization. However, OCTAVE requires a considerable expertise for appropriate application and is not applicable by a naive home user.

In 2001, Ross Anderson of Cambridge explained that economics can itself be a barrier to effective informations security design. [2] Anderson identified the need to align the incentives and the required investment in security. Net Trust aligns incentives with adoption in two ways. First, users control their own information rather than enriching a dot com. Second, Net Trust creates an open market for third parties. Thus, third parties may be created which are paid for by the user or supported by specialists groups. Currently third parties are all paid by producers, creating an inherent conflict of interest.

Later Gordon and Leob illustrated that information sharing is valuable, even when some participants are dishonest, or provide correct but incomplete information. [14] The focus of this work was on industry-specific ISACs with corporate participants. However, the game theoretic model is applicable to individuals who might err, lie or provide limited information, as implemented in Net Trust.

Further research verified that information sharing is not only valuable in and of itself but is also a complement to security investment. [12] This finding suggests that Net Trust will be not only valuable in itself, but may also increased overall user security awareness and investment. (Future research with Net Trust includes experimenting with Net Trust users to test for the existence of complementary investment, .e.g, using PGP or securing home networks.)

If Net Trust generates an increased awareness among individuals, it could also contribute to an increased awareness among those firms seeking those individuals as customers. Currently security investment is arguably inadequate. [13] Like firms, individuals suffer immediate costs and future risks from a loss of information integrity. In the case of firms, a security incident is associated with immediate loss of value. A study of capital market valuation of security incidents found that a firm loses more than 2% of its market value within two days of a publicized incident.

Net Trust also embeds past findings upon the economics of privacy. Why is it the case that the same individuals who express concerns about privacy will behave in a manner that systematically exposes their information? Without considering these findings, designs cannot appropriately embed a understanding of the human element.

Of course, first and foremost the privacy market needs reliable signals. “Protecting privacy“ is itself a vague promise. Privacy protecting software, as advertised, ranges from the private browsing enabled by the Anonymizer to the concentration of risk offered by Microsoft’s anti-phishing toolbar. [4]

Even when privacy can be defined and specified, e.g., through machine-readable P3P policies, a signaling problem remains. A model of the market with fluctuating numbers of reliable privacy-respecting merchants illustrates that the market will not necessarily reach an equilibrium where it is efficient for consumers to read privacy policies. Privacy policies are essentially first person assertions, where the merchant asks, "Trust me". Currently, there is no stable equilibrium under which consumers should read privacy policies. An external forcing function is required. Net Trust is designed to be such a forcing function by generating a trustworthy signal about, but not by, the merchant. [29]

Of course there is the argument that there is simply no market for privacy. However, there is certainly a market for privacy off-line. Products from simple window shades (with unarguably limited aesthetic appeal) to locking mailboxes thrive in the physical realm. Observations of the physical and virtual markets for products providing privacy suggests that, "when privacy is offered in a clear and comprehensible manner, it sells". [26] Net Trust makes privacy and security clear and comprehensible.

Privacy can be good or bad for individuals, if the information obtained by others' is used to lower prices or to extend privileges. In particular, the opposite of privacy in the market is not necessarily information; the opposite of privacy is price discrimination. In markets where there is zero marginal cost (e.g., information markets) firms must be able to extract consumer surplus by price discrimination. This means that the firms cannot charge what they pay, at the margin, but must charge what the consumer is willing to pay. What are privacy violations to the consumer may be necessary pricing data to the merchant. [21] Accurate signaling information, while useful for the market may not be in the interest of firms and thus never receive support. Therefore peer production of information about merchant reliability is arguably necessary.

Indeed, individual rejection of security information may itself be rational. When information security means ensuring that the end user has no place to hide his or her own information, or when security is implemented to exert detailed control over employees individuals rightly seek to subvert the control. Security is often built with perverse incentives. Privacy and security are constructed to be opposites instead of complements in controlling information. Rejection of security is, in some cases, strictly rational. [25] Net Trust has been designed to be incentive-aware and to align with the incentives of the end user, not the firm.

After a highly publicized security breach at the data broker Choicepoint, the Choicepoint Chief Security Officer claimed, "Look, I'm the chief information security officer. Fraud doesn't relate to me." [17] If the CSO of Choicepoint finds addressing fraud beyond his reach, imagine the tenuous situation of the average computer user. Net Trust allows these users to help each other by using peer production of signals.

2.2 Peer Signaling in Resource Allocation

The literature on peer production was popularized by open code and p2p networks. Peer production has been found to have many advantages over firm-based production. Peer production effects the modularity, granularity, and cost of integration of a good produced: it shifts the distribution of production costs to those most able and willing to bear them. [3]

The peer production of information in Net Trust is highly modular. The granularity of Net Trust's implicit rating system is the URL. The use of social networks to group people affords Net Trust many advantages. Although limiting the radius of contacts from which information can be gleaned vastly constrains the total quantity of information, the smaller egocentric network of each individual is composed of trusted individuals. (These individuals are trusted in the social sense, not trusted in the cryptographic provable sense.) Users will trust themselves not to invite a malicious actor into their personal networks; they are less likely to trust the judgment of friends of friends. Trust may be transitive, as the commonly cited BBK model indicates, if so it's with a finite radius. [27] A smaller network, both numerically and socially, has a smaller chance of containing a malicious node. The BBK model is predicated on numerical weights of how much each node trusts their neighbors. The Net Trust model requires no such calculation; rather personal selection is a Boolean indication of trust.

A decrease in free riding is a possible outcome of a small network. With a small social network the incentive to cheat or free ride is less severe than in a larger anonymous system. Members of a social network will have less benefit from unfriendly behavior since there are fewer known people to damage (unlike with a centralized recommender) and they share social ties with at least their immediate neighbors. The use of implicit data means that free riding, simply having the system passively obtain information from others without generating any information, will still be an issue in Net Trust. Such behavior should be encouraged, when the alternative is non-use.

There are practical advantages to a small network. The number of communication links scales as a square of the total network size [24], and smaller network decreases overall traffic. In addition to scaling up more easily, the fact only a limited number of individuals are needed for the system to work means that Net Trust can grow without waiting for critical mass.

Finally, early usability test (Section 4) and the theoretical discussion above indicate that individuals are more interested in signals from their immediate social network than in global systems. Social networks may differ in terms of their perception of what is a legitimate or desirable site. For example, few who shop at Prada are likely to embrace Kmart's shoe sales; while few who buy shoes at Kmart will find Prada's pricing reasonable. Quality does not need to be a global property, as long as it has a local meaning inside a given segment of the social network.

Finally, it has been shown that the presence of a small, persistent application dedicated to a specific purpose will raise user awareness and consciousness about that concept, even if use of that application is minimal. [9] This echoes the finding in [13] that suggests security information and investment are complements.

The Net Trust system was designed first and foremost to address malicious, fraudulent or masquerading websites. The basic model, however, is very extensible. Net Trust incorporates implicit behavior-driven ratings, explicit individual recommendations, and personally selected trusted parties. Conceptually, this user model would be useful for any resource of unknown quality when the resource quality is static. That is, a bad resource cannot strategically behave as a good resource some fraction of the time. For implicit information to work, the distribution of resources must not be independent of the distribution of users across the system. That is, for Net Trust to work users in a self-selected social network should be more alike than a random group of strangers. In fact, this correlation does not have to be very large if certain assumptions are made about the social network structure.

Indeed, there are many situations when the above conditions hold, and sharing with trusted social contacts is superior to sharing a group of strangers. With the implicit rating, social network-driven recommender systems could be used to address the additional generic quality problems. For example, interdisciplinary researchers cannot always ascertain which research journals are superior in fields that are not their disciplinary home. Net Trust could be used to track either publishing or reading habits of an academic peer group, allowing each member to gauge the relative importance of a journal based on readership. Net Trust may add more value than indication of phishing sites. That increased value will make Net Trust more usable (as it is perceived as worthwhile) and more used (as individuals integrate its signals into their decision-making).

Peer production makes achievable production that cannot be done with centralized capital. The context information that can be re-embedded by constraining production to known, trusted and similar peers is very powerful.

3 Integrating Privacy-Enhanced Signaling into Browsing

Net Trust is radical and exploratory is that it was based first on economic theory and second on a set of user tests. This paper has described a new type of application, conceived of in economics, modeled in theory, and tested in the laboratory.

Net Trust is based in a toolbar interface and a p2p back end. Some toolbars target specific threats for example, phishing. Spoofguard is one of the toolbars using real-time characteristics of the phishing sites themselves; for example, links, images, lack of a SSL certificate, or misdirection in the links. These variables are all under the control of the malicious agent. In contrast, Net Trust uses features that are not under the control of the malicious agent: user social network, user history and the history of the user's social network. In addition, the Net Trust toolbar takes advantage of a characteristic of phishing sites to prevent one phishing victim from misdirecting others - the temporal history of phishing sites. Phishing sites go up, are identified, and are taken down. Phishing sites do not stay up over long periods of time. Therefore the impermanence of phishing sites is integrated into the reputation system as described below.

Net Trust is a toolbar plug-in for a web browser. The toolbar will be the main source of information to the user. The application has six primary components: the social network, the reputation system, the third parties, the interface, and the data distribution network.

Net Trust enables integration of information from trusted parties. These are not trusted third parties, as is the tradition in security and cryptographic systems. The removal of the word "third" in the traditional trusted third party construct indicates that the individual makes the final trust decision, not the trusted party. There is no root that determines which parties are trusted. In trusted third party systems, the browser manufacturer, employer or other third party often determines who is a trusted party. In Net Trust certificates are self-signed, and users select trusted information providers. The Net Trust user, not the distributor or developer, makes the final determination of which parties are trusted. Compare this with SSL or Active X, where the user is provided a mechanism to place trust in a third party. After the initial selection of the third party, the user's decision is implemented by technical fiat by accepting those the third party validates.

Net Trust integrates social network information. Individuals may have multiple social networks: home, family, hobby, political or religious. Regardless of the level of overlap, information from one social network may be inappropriate for another social network. Not only do people share different information with different people but also different social networks are associated with different levels of trust. [10] For example, professional colleagues can have much to offer in terms of evaluation of professional websites, but professional interactions are not characterized by the same level of openness as family. Professional networks are not systematically used to request intimate or religious information. Because of these differences, overlapping contexts can cause a breach in privacy. In order to support the construction of boundaries between one person's various roles, the application allows a user to have multiple identities (e.g., pseudonyms) coupled with multiple social networks. Pseudonyms engage in disparate social networks. Members of that social network are called "buddies" both to indicate the similarity to other online connections and to indicate that the standard for inclusion may vary. "Buddy" is also sufficiently vague to communicate that there is no standard for strength of the network tie. When a user leaves, departs, or clicks out of a website the URL is associated with the pseudonyms visible in the toolbar. Choosing to associate a site upon departure instead of arrival allows users to make informed selections of websites. Once a website has been identified as associated with a pseudonym (in the figure shown the pseudonym is " Alex Work") the user no longer has to select that identity when visiting the associated website. If Alex is in work mode, and then visits a site she has identified as associated with the Alex@home pseudonym, Net Trust will change pseudonyms at the site. If Alex wants to share a site across pseudonyms, he has to make a non-zero effort (holding down a control key) to add additional pseudonyms to a site already associated with one pseudonym. Therefore after a website has been associated with a pseudonym all future visits correspond to that pseudonym, regardless of the website selection at the time of site entry. Thus individuals have to make pseudonym choices only on new websites. Presumably individuals will select a default pseudonym, possibly different pseudonyms for different machines, e.g., at work or home.

3.1 The Buddy List

An essential component of this application is the re-embedding of a user's existing social network into their online browsing experience. In brick and mortar commerce, physical location is inherently associated with social network, as exemplified by the corner store, regulars at businesses, and local meeting places. Net Trust uses social networks to capture virtual locality information in a manner analogous to physical information. Net Trust implements social networks by requiring explicit interaction of the user. The Net Trust user creates a " buddy list" containing the social network associated with a pseudonym. Using the Net Trust invitation mechanism, a user sends a request to a buddy asking for authorization to add them to the user's buddy list. Once the buddy approves the request, the user can place the buddy in the social network defined by the appropriate pseudonym. Social networks can be presented for user consideration from importing IM lists, email sent lists, or pre-existing social network tools such as Orkut, Friendster, Face Book, or LinkedIn. Net Trust requires that the individual issuing the invitation to his or her buddy know the email or IM of that buddy. The invitation includes file location information that must be integrated into the distributed file system. The following description identifies a file name as the minimal adequate locator.

Consider a Net Trust user named Alice who has as an associate a person named Bob. Unlike standard cryptographic protocol descriptions, we assume that Bob and Alice have established a virtual social history. Before inviting anyone to a network Alice creates a pseudonym. Once the pseudonym is created she creates a set of asymmetric keys, public and private. For simplicity, call the pseudonym Alicework. The private key allows Alice to confirm that any message from Alicework came from Alicework to anyone with the corresponding public key. Alice sends an invitation with a nonce to Bob. The nonce prevents replay attacks and ensures freshness. The public key prevents anyone else from associating themselves with Alice's pseudonyms after the initial introduction. The public key is not published. The example can only continue if Bob agrees to join the system. Because Bob joins the system, Alice will share Alicework's history with Bob's chosen pseudonym. The history-based reputation information is contained in a file or feed that is identified by a 128 bit random number. The feed or file will not include personally identifiable information. Since Alice initiated the invitation, she sends Bob her file locator and a key used to sign the file. Then Bob will send his file locator and a key used to sign his feed. Part of Bob's choice includes filling out information about his affiliation with Alice - her name and his corresponding pseudonym, as well as a review date for her inclusion. Thus interaction is designed to cause joining a stranger's network to cause some cognitive dissonance by demanding unknown information in order to consummate the introduction. Indeed, social network sizes are fixed so that position in someone's social network has value. Were social networks expandable to the thousands, then choosing to join someone's network would be the default. Limiting the number of possible participants in a pseudonym is designed to decrease the likelihood that a stranger will be able to join. (After distribution of Net Trust, we hope to implement experiments to test how likely Net Trust users are to accept a stranger to their social networks.)

After this introduction Alice and Bob update each other's own local reputation-based signals by sending out information. Alice and Bob update their own ratings by periodically downloading each other's published files. The files, designated " filename" , include URLs, ratings, and dates. Bob's ratings are then integrated into Alice's toolbar as Bob's opinions of sites, with Alice's client reading Bob's file and Bob's client reading Alice's file. The data are public and signed, but not linked to any identity excluding via traffic analysis. (The importance of traffic analysis underscores the use of Tor in this system, as described in Section 5.)

In the proposed initial instantiation of Net Trust, different individual's opinions are not to be differently weighed. Segregating individuals into social networks creates implicit weighting. Some systems assume that individuals should be provided with different trust weights because some contribute more than others. [23] In contrast, our system allows the user to evaluate his or her own context and weigh based on the provision of the information. While the proverbial grandparent might not be as apt at discriminating between legitimate and malicious sites as a computer savvy co-worker, she may have extensive knowledge of hunger-based charities from volunteer work or detailed knowledge of travel locales from personal experience. Therefore, our initial design asserts that there is no single trust weight for an individual across all contexts. By simply hitting the icon of " people" as seen near the left in the image of the toolbar, the user will see an enlarged view of their social network and pertinent browsing statistics. User-selected icons are displayed for ease of identification and personalization. Net Trust also allows for the addition of third parties who make assertions about trust as shown in the toolbar above. Centralized trusted parties provide these ratings. They are called " broadcasters" in this model to emphasize that they distribute but do not collect information. While buddies share information by both sending information and obtaining regular updates, broadcasters only distribute information. Broadcasters use a certificate-based system to distribute their own files, with Boolean ratings. Such lists of " good" and " bad" sites are sometimes called white and black lists or green and red lists. These lists are stored and searched locally to prevent the need for the Net Trust user to send queries that indicate their browsing habits. Requiring a web query for searching would create a record of the client's travels across the web, as with Page Rank records on the Google toolbar and Microsoft anti-phishing toolbar. Broadcaster's ratings are shown as positive with a happy face, negative as a yuck face, and no opinion as blank. Early user test indicated that users could misunderstand a neutral face as a positive or negative assertion. Indeed, early user test found that signals less blunt than smiling and yucking faces were confusing. The default on a URL that is not included in the ratings provided by the broadcaster, the default is to have nothing displayed.

Net Trust users will be able to select their own broadcasters. Individuals that can be fooled into downloading false green lists can be undermined by this system. To mitigate the possible harm there is a maximum lifetime for any green list. Broadcasters can be removed, but it is not possible for an attacker to replace one broadcaster with another even if the first one has been removed. (Being able to override that feature requires that an attacker have write permission on a users' drive. At that point, user trust of websites becomes a negligible measure of security.) Since the broadcasters provide important information, like any other trust vector, subversion of that trust can cause harm. However, since broadcasters only inform trust decisions the harm is limited and if there is bad information the source of the bad information can be detected by the user. Compare this with Active X or the addition of trusted certificate authorities, which alter authorization and thus access on the user's machine. In those cases malicious action from code cannot be determined during regular use by the user. Both of these systems and broadcasters embed expiration dates.

The security of this system depends on the ability to identify network participants reliably and prevent leakage of the key used to share history. If attackers can rewrite histories the system is a net loss in security terms. There is no universal identity infrastructure on which this system can depend. Invitations are issued by email and email: identity authentication is arguably tenuous. Social viruses have long utilized the lack of authentication in email to increase the likelihood of victims taking the necessary action to launch the virus. However, by requiring the response, this mechanism cannot be subverted by mere deception in the " From" field.

The security policy is one that is based more on economics than on traditional security. The Net Trust system as modeled using economics assumptions will create value for users and increase the difficulty of certain types of financially motivated attacks. The next section, Section 3.2, describes the reputation system.

3.2 The Reputation System

The current reputation system has been modeled is described above. An initial visit will log the web site on the basis of domain names. This will create a rating of 1. That rating will decay uniformly so that if the site is not visited again the rating goes down to 0.5. Each time the website is visited the rating will double, to a maximum of n. Currently, n is set to 10. When a site is explicitly rated, the explicit rating remains without decay. Only explicit user action can change a rating based on previous explicit user action. In the case of a negative rating, the social network window shows a large red bar connecting the user to the site. The lowest implicit rating is zero. for each website that is associated with a pseudonym there is one of the two possible records: either this with explicit rates

1. url, date initail visit, number of visits, last visit¿

or this with explicit rating

2. url, explicit rating.

To restate the reputation in mathematical terms the reputation calculation mechanisms is as follows, with the rating value is R_w.

- For sites with no visits or for a visit less than t_0 previously $R_w = 0$
- For one visit, more than t_0 but less than t_d hours ago $R_w = 1$
- For m visits with a last visit having occurred at $t - t_m > t_d$ $R_w = \min\{n, \max\{m/2, me^{-c|t-t_m|}\}\}$

Recall from the description above that R_w is the reputation of the website, with a maximum value of 10 in our current model, m is the number of visits, t_m is the date of the most recent visit, t_d is the decay delay parameter, and t is the current date. c is simply a normalizing constant, to prevent too rapid a decay in the reputation after t_d.

Current phishing statistics suggests a value of t_0 of not less than twenty four hours; however, this may change over time. One system design question is if users should be able to easily write t_0 or t_d; i.e., if the system should be designed to allow an easy update if new attack with a greater temporal signature is created. If the value of t_0 it is too low then attack sites could change victims to supporters too quickly. Thus being able to increase the value offers the opportunity for a more secure mechanism. However, the value to alter t_0 can itself become a security risk, as a phisher could convince users to set $t_0 = 0$.

To summarize the reputation system, a single visit yields the initial rating of 1 after some delay. The delay time prevents those who are phished early from becoming agents of infection, and supporting later phishing attacks. Then as the number of visits increases the score itself increases in value. The least value for a visited site that has not been manually rated is zero. The greatest reputationvalue for any site is 10. The least reputation value of any site is -10.

Current phishing statistics suggests a minimum value of t_0 of 24 hours, we will use up to 168 hrs. One system design question is if users should be able to easily write t_0; or if the system should be designed to allow an easy update if a new attack with a different temporal signature is created. If the value of t_0 is too low then attack sites could create reputation quickly and then use Net Trust to effectively send powerful false signals. The ability to alter t_0 can create increase the value of the signal, or itself become security risk. . Consider a phishing website. For a phishing website any broadcasters will label the site as bad or neutral. No member of the social network will have ever visited the site. While this may not deter someone from entering a site to shop for something unusual, it is an extremely unlikely outcome for a local bank, Pay Pal, or eBay. In order to increase the efficacy of the toolbar against phishing in particular, one element of the project entails bootstrapping all banking sites. Those websites that are operated by FDIC-insured entities are identified by a positive signal (a smiley face). Those websites that are not FDIC institutions are identified by a negative signal (a yuck face). The icons are shown and described in the previous section. In addition, bootstrapping information can be provided by a compendium of shared bookmarks (Give-A-Link) or SiteAdvisor. PhishGuard generates a list of phishing sites and could be integrated into Net Trust. PhishGuard uses peer production of information by asking people to submit phishing sites, but provides no privacy or other feedback.

Without the inclusion of the FDIC listing then the Net Trust toolbar has a failure similar to many security mechanisms where the user is forced to look for what is not there. Seals function if they are not only noted as present but also noticed when missing. The lock icon on SSL is replaced with red icon, but the user must notice that the lock is missing. In email, eBay messages include a header indicating that only messages with the eBay header are to be trusted. Obviously faked emails do not include a flag to indicate that the expected header is missing. Trust seals are easy to copy. SSL-secured phishing attacks have already occurred. What is truly missing is valid economic signals for resource identification on the web.

The long-term efficacy of the reputation system depends upon how similar social networks are in terms of browsing. Do friends visit the same websites? Do coworkers visit the same website? For example, in the most general reputation system where every user had a given chance of seeing any one page from a known distribution, could correctly judge a bad resource as such with probability p, and would mislabel it with the corresponding probability 1-p a decision rule could trivially be derived. However, that information is not only unavailable for small social networks but the data are also generally unavailable. This research requires that Net Trust be completed and have a group of users. Using this reputation system and with the assumption of different degrees of homophily the user modeling as described above indicates that Net Trust would provide a high degree of value in identification of sites. Homophily means that people who are in a social network have similar browsing habits. Users are not uniformly distributed across the low-traffic sites on the web. Some sites of are interest only to a small population, such as members of a course at a university, or members of one school or office.

Given that the ideal mechanism cannot be known because social network homophily is not known, the implementation is based on user modeling. Recall that the user modeling indicates that this toolbar will enable a significant increase in the ability of end users to discriminate between resource types. The model indicates that the inclusion of bootstrapping information will dramatically increase the ability of end users to discriminate. The modeling of the reputation system indicates that the system as proposed will be valuable in assisting users in distinguishing between good and bad resources.

4 Usability Study Results

Net Trust is only useful to the extent that it is usable. Thus Net Trust began with user testing. Twenty-five Indiana University graduate and undergraduate students participated in the first usability study of Net Trust and fifty in the second. The students were from the School of Informatics. Initially, the participants of the usability study were asked to spend a few minutes investigating three websites. The websites were fabricated especially for the purpose of a usability study, and therefore controlled for content and interface. (The similarity of the three websites was tested at Loyola Marymount before the Net Trust experiments.) The participants were asked to indicate if they would trust the sites with their personal identifiable information, including some financial information. In the first test, the toolbar was enabled on in the browser and the participants were instructed to visit each of the three websites again and complete one toolbar task on each site. The tasks included rating a site, adding and removing buddies, as well as switching between buddy and network view. The survey had been previously validated with two tests of undergraduates. For the Net Trust usability test, the toolbars were seeded with reputation information. In the second test, users were separated into those with and without the toolbars.

Afterward examining the websites, the participants were prompted to indicate their trust of the three websites taking into account the information provided by the toolbar. For the first two websites, the toolbar showed a large number of “ buddies“ visiting the site, 6 out of 10 for website 1 and 8 out of 10 for website 2, respectively, as well as positive or neutral ratings for the broadcasters. The last website showed only 2 out of 10 friends visiting the site and negative or neutral rating from the broadcasters. The toolbar significantly increased the propensity to trust a website. The results demonstrate that the toolbar is successful in providing a signal of trust towards a website. Even when the toolbar showed a significant amount of negative ratings, such as in website 3, the fact that a website had been previously visited by members of a social network increased the propensity to trust. This finding is validated by the examination of trust mechanisms described earlier in the paper that argued that social networks are a most powerful mechanism for enabling trust.

4.1 Privacy Considerations and Anonymity Models

The critical observation of the privacy of Net Trust as it is proposed here is that the end user has control over his or her own information. Privacy can be violated when a user makes bad decisions about with whom to share information. However, the system does not concentrate data nor compel disclosure. There is a default pseudonym in the system that is shared with no other party (private) and another that collects no information at all (logout).

Net Trust is designed to ensure privacy, in that the users can share selected information; withhold information; and can control with whom they share information. This system shares web browsing information in a closed network of peers. In contrast, recall Furl and Del.icio.us. Both are designed to leverage the observation that each user has a unique view of the web informed by their own history and the history of others. In both systems there is significant centralized storage of user browsing and no explicit mechanism for user pseudonyms. Neither of these systems uses the developments in social network systems beyond simple collaborative filtering. Individuals do not select their own peer group. As a result information can be inappropriate and in some cases data are highly polarized; for example, a search for “George W Bush“ on Del.icio.us yields images of both a president and of various chimps. Del.icio.us and Furl do have a commonality with Net Trust in that they leverage the similarity of browsing patterns.

Net Trust also differs other social browsing mechanisms in that identity is no intended to universal. There can be many Bobs as long as there is only one Bob in any particular social network. Effectively, identities are used as handles or buddy names to create a virtual implementation of a pre-existing social network. Identity construction assumes a previous context, so that meaning is derived from the name and context. Each person can construct as many pseudonyms as he or she desires, where each pseudonym corresponds to a distinct user-selected social network.

Net Trust is designed to have three default pseudonyms: userhome, userwork, and private. Websites visited under the private pseudonym are never distributed in the Net Trust data structures. There is an argument for keeping a “private“ list stored. The advantage is that users can inform their own personal browsing. The disadvantage is that the user may want nothing recorded. Our solution is to have a private pseudonym and an option to logout from the system. “Logout“ is not considered a pseudonym.

In all cases, if a user is logged in under a certain pseudonym, her website activity will only be shared with the social network associated with that pseudonym, not with any other networks that might exist under different pseudonyms. The user may also edit by hand the list of sites that is shared with any particular social network. Subsequently, a user's online activity is only used to inform the buddy view in other buddies' handpicked views. Becoming and offering oneself as a broadcaster requires downloading additional software, as well as publishing a public key. The interface for broadcasters in our design accepts only single entry URLs and requires notations for each URL entered. Our design is directed at preventing anyone from becoming a broadcaster by accident. There is no implicit rating mechanism for broadcasters.

5 Future Research: Distribution of Files on the Network

The server implementation of Net Trust, where the feeds are centralized in one or two servers, enables traffic analysis. The fact that it provides the ability to obtain aggregate information across multiple feeds (which do not have personally identifiable information) can be a benefit. While we acknowledge the potential benefit of aggregate information, addressing traffic analysis is the next step planned in this research.

Here we discuss possible systems for distributing content. The requirements are modest. The lookup protocol needs to map a filename to a request. The storage protocol must store, cache and retrieve data. Overwriting data is an acceptable threat, as long as the client reliably detects the loss of integrity. Authentication of data occurs in the client. The file system needs to defeat traffic analysis and should be non-immutable. The privacy requirements are first and foremost that no one should be able to discover another person's social network. Second, each person should be able to share information reliably in his or her social network and not others. Within this we recognize that correlation of identity to filename identifiers will inevitably occur in some cases. However, identification of an individual should not negate that person's ability to control his or her own data. Here we list some of the options for the distribution and storage of data.

Chord is a symmetric option. The Chord lookup mechanism is robust in the face of frequent node failures and rejoins, which describes the Net Trust system. Chord protocol involves maps a key to a node using consistent hashing. Each Chord node needs routing information about only a few other nodes. While for a single Chord node traffic analysis with other known nodes is possible, creating a level of protection from traffic protection from third party observers.

Freenet is a censorship-resistant peer-to-peer network. Freenet uses key based routing, which is similar to distributed hash tables, to locate peer data. It is encrypted, replicated and anonymous. Each node maintains a data store containing documents associated with keys and a routing table associating nodes with records of their performance retrieving keys. Freenet by design does not allow easy replacement of older files, as it is designed for immutable files. Net Trust will have frequently updated files, so the match between Freenet and Net Trust is not particularly good.

Mnet is a searchable, encrypted data store. It is the successor to Mojo Nation. It doesn't provide full anonymity. It offers limited deniability, in that one doesn't have knowledge of what is being stored. It is censorship resistant as files are split across different nodes. As with Freenet, content loss is a larger threat than censorship for Net trust.

Tapestry provides location-independent routing based on filenames. In that way, mobile individuals with consistent random file names may be well served. This may make it easier for individuals to access their own histories and social networks from remote locations. Tapestry is self-organizing, fault resilient and load balancing. However, Tapestry does not defeat the problem of traffic analysis.

Pastry is decentralized, scalable, and self-organizing. In Pastry, object replication decreases the risk of traffic analysis at the cost of loss social network updates. The Pastry assignment of a 128-bit nodeID is uniformly random, so the system would select the user filename as opposed to Net trust generation. Pastry could be compatible with Net Trust but would require significant alterations.

Publius is not fault tolerant and is not searchable. It has good anonymity and has automatic caching and replication. However, file updating can be expensive as Publius was designed to be censorship-resistant as opposed to easily updated.

Mnemosyne is a distributed and secure P2P backup system. It creates several copies of a user's file and backs them up in other machines connected. It is anonymous and all files are encrypted. Mnemosyne is to be a feasible alternative but does not appear active and available.

Tor is the next generation of onion routing. Tor has high encryption costs, per message. However, the sending and retrieving of buddy signals is sporadic, not constant so the delay should be acceptable. Tor solves the issue of traffic analysis. Tor requires more overhead during initiation for route identification. However, if file locations stay constant then the overhead may be distributed over the multiple writes envisioned for Net Trust user data. Tor has a working infrastructure, a published API, and can be integrated into Net Trust most reliably. Tor hidden services can distribute information to each social group and simultaneously prevent any person from discovering another's social group. To work with Net Trust, Tor must be integrated with a data storage mechanism.

6 Conclusions

Privacy and security markets in general suffer from a lack of signaling. This decreases the overall demand for privacy and security information and products, as neither is necessarily trustworthy. One reason there is not reliable signaling is that the incentives for producers of trusted information is to sell that information. In a adaptation of Gresham's law, bad security can drive out good when both demand the same price. Indeed, high levels of Internet fraud make individuals less likely to interact with trustworthy sites.

In order to provide reliable signals and enable users to make informed trust decisions we are developing and testing a new type of security application, Net Trust. Net Trust is grounded in economic research and tested for human interaction. Net Trust integrates information from multiple user-selected sources to create a single easy-to-evaluate, contextually appropriate signal.

References

[1] A. Acquisti and Jens Grossklags. Privacy attitudes and privacy behavior. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, chapter 13, pages 165–178. Springer, New York, NY, 2004.

[2] R. Anderson. Why information security is hard: an economic perspective. In *ACSAC 01: Proceedings of the 17th Annual Computer Security Applications Conference*, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[3] Y. Benkler. Coase's penguin, or linux and the nature of the firm. *Yale Law Journal*, 112, 2002.

[4] L. Jean Camp and Carlos Osorio. Privacy enhancing technologies for internet commerce. In *Trust in the Network Economy*. Springer-Verlag, New York, NY, 2003.

[5] L. Jean Camp and Catherine Wolfram. Pricing security. In *Proceedings of the CERT Information Survivability Workshop*, page 3139. CERT, Pittsburg, PA, 2000.

[6] Huseyin Cavusoglu. Economics of it security management. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, page 7183. Springer, New York, NY, 2004.

[7] Federal Trade Commission. Ftc releases top 10 consumer complain categories for 2004. Technical report, Federal Trade Commission, Washington, DC, February 2005.

[8] Federal Trade Commission. Ftc releases top 10 consumer complain categories for 2004. Technical report, Federal Trade Commission, February 2005.

[9] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. *Use of a P3P user agent by early adopters*, pages 1–10. ACM Press, New York, NY, USA, 2002.

[10] J. Donath and D. Boyd. Public displays of connection. *BT Technology Journal*, 22(4), October 2004.

[11] C. Dwork and M. Naor. Pricing via processing, or, combating junk mail. In *Advances in Cryptology CRYPTO92, Lecture Notes in Computer Science*, page 139147. Springer, 1993.

[12] Esther Gal/-Or and Anindya Ghose. The economic consequences of sharing security information. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, chapter 8, pages 95–105. Springer, New York, NY, 2004.

[13] L. A. Gordon and Martin Leob. The economics of information security investment. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, chapter 9, page 106123. Springer, New York, NY, 2004.

[14] Lawrence A. Gordon. An economics perspective on the sharing of information related to security breaches: Concepts and empirical evidence. In *Workshop on the Economics of Information Security*. UC Berkeley, Berkeley, CA, USA, May 2002.

[15] S. Grabner-Kraeuter. The role of consumers' trust in online-shopping. *Journal of Business Ethics*, 39, August 2002.

[16] R. Kalakota and A.B. Whinston. *Electronic Commerce*. Addison Wesley, Boston, MA, 1997.

[17] C. Koch. The five most shocking things about the choicepoint debacle. *CSO*, 2005.

[18] Ben Laurie and Richard Clayton. Proof-of-work proves not to work. In *Third Workshop on the Economics of Information Security*, Minneapolis, MN, USA, June 2004.

[19] Thomas A Longstaff, Rich Pethia, C Chittister, and Y Y Haimes. Are we forgetting the risks of information technology. *IEEE Computer*, pages 43–52, 2001.

[20] H. Nissenbaum. Securing trust online: Wisdom or oxymoron. *Boston University Law Review*, 81(3):635–664, June 2001.

[21] A. Odlyzko. Privacy, economics and price discrimination on the internet. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, page 187212. Springer, New York, NY, 2004.

[22] J. Riegelsberger and A. Sasse. Trustbuilders and trustbusters. In *IFIP Conference Proc*, volume 20, 2001.

[23] Paul Syverson R.oger Dingledine, Nick Mathewson. Reputation in p2p anonymity systems. In *Workshop on Economics of p2p Systems*. ACM Press, New York, NY, USA, 2003.

[24] G. Prasanna S. Marti and H. Garcia-Molina. Dht routing using social links. In *3rd International Workshop on Peer-to-Peer Systems*, 2004.

[25] M. Sandrini and F. Cerbine. We want security but we hate it. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, page 213224. Springer, New York, NY, 2004.

[26] A. Shostack and Paul Sylverson. What price privacy? In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, page 129142. Springer, New York, NY, 2004.

[27] M. Borcherding T. Beth and Klein. Valuation of trust in open networks. In D. Gollman, editor, *Computer Security — ESORICS '94 Lecture Notes in Computer Science*, pages 3–18. Springer-Verlag, New York, NY, 1994.

[28] Pew Charitable Trusts. Getting serious online, 2002.

[29] T. Vila, R. Greenstadt, and D. Molnar. Why we cannot be bothered to read privacy policies. In L Jean Camp and Stephen Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, page 143154. Springer, New York, NY, 2004.

[30] Y. Wang, D. Beck, Z. Jiang, R. Roussev, C. Verbowski, S. Chen, and S. King. Automated web patrol with strider honeymonkeys: Finding web sites that exploit browser vulnerabilities. In *Proc. Network and Distributed System Security NDSS Symposium*. ISOC, 2006.