Someone to Watch Over Me

Heather Richter Lipford
Department of Software and Information Systems
University of North Carolina at Charlotte
9201 University City Blvd., Charlotte, NC 28223
+1 704-687-8376
Heather.Lipford@uncc.edu

Mary Ellen Zurko
Cisco Systems
1414 Massachusetts Avenue
Boxborough, MA 01719
+1 978 936-6396
mzurko@cisco.com

ABSTRACT
Traditional security mechanisms are part of a larger socio-technical system involving the people and organizations that use them. Yet, those security mechanisms rarely take this social context and social processes into account. In this paper we propose to make security more social, by integrating community oversight into security mechanisms. Like a neighborhood watch, community oversight can provide additional information as more people are able to detect anomalies and problems, as well as foster greater awareness and social norms of security-related behaviors. We describe this new paradigm, several scenarios of use, and the sets of issues involved in implementing this approach.

Categories and Subject Descriptors
K.6.5: [**Management of Computing and Information Systems**]: Security and Protection.

General Terms
Security, Human Factors.

Keywords
Security, social influence, community, oversight.

1. INTRODUCTION
Social processes govern much of our lives, from how we behave around others, to how we make decisions, and even how we protect ourselves from harm. There are many examples of community or interpersonal oversight which provide some form of security or safety in the physical world. For example, employees are expected to notice people without appropriate badges and to disallow tailgating at entrances. Airport customers are asked to report suspicious activities and packages. Neighbors alert each other when garage doors or windows are accidentally left open. Neighborhood Watch programs come in many types, and have been associated with a decrease in crime [20] through increased surveillance and helping establish norms of behavior and intervention among residents. Yet, security solutions often rely on solely technical mechanisms, not taking advantage of the social and organizational context of the users.

NSPW'12, September 18–21, 2012, Bertinoro, Italy.

In this paper, we propose and discuss community oversight as a new paradigm within security systems. Oversight in security systems has long been problematic (Quis custodiet ipsos custodes?). Most often it is conflated with administrative type privileges, or in more careful contexts, separated out as a specific and orthogonal role of its own. We consider how oversight through our interpersonal connectedness and exposure through our current computer systems might be used to provide new forms of security, to augment existing security, or to replace existing security mechanisms. Taking inspiration from the neighborhood watch, we look for ways that a user's friends or community might watch over her, or how a community might watch over itself, and the security benefits that can provide. And similar to a neighborhood watch, we emphasize both the added information that can be gathered from community oversight of security and the impact on security-related social norms and behaviors that will occur. Thus, the paradigm is less about oversight that directly catches malfeasance, and more about creating cultures and norms around behaviors that ward off problems and exposures, and an awareness of what those behaviors are.

A key problem in information security in organizations is that employees may not comply with security procedures [19]. Yet enforcement of those policies is challenging due to the expense and reliability of auditing individuals [19][33]. Instead, organizations need to institute a culture of compliance towards information security policies, increasing awareness and incentives for desired security behaviors [18][33]. For example, Herath and Rao have found that that the intent to comply with security procedures is significantly influenced by employees' perceptions of norms within the organization [19], and their perceived certainty of detection. Pieters and Coles-Kemp [30] point out that conflicts between cultural norms and information security policy must be recognized and reduced at the design phase of security systems. This research motivates our approach by emphasizing the importance of a user's social context upon security behaviors. We expand upon these results by exploring integrating community mechanisms that may influence compliance directly within the security technologies themselves, potentially supporting a wide variety of communities and security decisions.

Examples of informal and useful group oversight of computer use are hardly new. In the early 80's, real time shared use of a computer was common. For example, one of the authors, Zurko, belonged to a group of professionals who shared a VAX for their work which did not require a dedicated machine. Because shared CPU was a critical resource, the team had a rule that computer games could only be played during the lunch hour. Members of the group running games outside of that time (and during work hours) were quickly caught and stopped by team members, who

had the permissions and knowledge to see what was running on the computer and halt it. Group oversight was effective in this case, in part, because the business rules were directly and obviously aligned with benefits to the group members themselves. As such examples demonstrate, communities of people are able to employ social processes to enforce desired behaviors when information and actions needed for such oversight are visible to them. We believe security systems should explicitly provide such information and methods to enable these social processes.

We see a number of ways that structured or unstructured community oversight might be used to benefit security. Most obviously from many existing examples, it can be a method to watch for anomalous behavior or irregular activity, or for spotting attacks and problems. It could be a method to bring about a culture of compliance to security policies, impacting or noting both mistakes and conscious violations. Taken even further, this approach can enhance the ability for communities to control security through the generation of useful social norms; using social influence to spread and enforce good security information and practices. Questions of awareness, education, training, and philosophical alignment arise in many discussions of compliance to security policies. Choosing what community members might see or watch has potential for impact on all of those aspects. We believe that by directly supporting security-related community oversight, we can complement traditional security by addressing the weaknesses in such technologies, by influencing the decisions and behaviors of people to better utilize existing technologies and follow security policies, and to apply collective human judgment towards noticing anomalous or potentially dangerous behaviors.

In this paper, we present the general idea of community involvement in security-related oversight and its potential benefits and research issues. Rather than present one particular solution in detail, we present this idea to seed discussion and future research in this area. We demonstrate its generality with three broad scenarios of use in different kinds of communities: organizations, online social networks, and families. We then discuss the variety of issues regarding the design and implementation of such a mechanism, as well as discuss potential research needs and limitations.

2. BACKGROUND

There are currently many ad-hoc methods where communities of people form to help each other with security-related issues. Countless discussion forums and websites exist for people seeking out advice and updates on security-related information. Security professionals often rely on email lists and websites for advance warning of attacks and exploits [22], and both professionals and average consumers seek advice on specific settings and on how to recover from problems. Another example is Web of Trust, a crowd-sourced web rating system, providing feedback from a community of users about the trustworthiness of websites [8][37]. While these methods provide valuable information, they are not integrated in the security system itself, requiring the user to go seek out and understand the information explicitly, reducing their impact and use.

Several security solutions also have similar motivation, but without as much active participation of users. For example, the Perspectives and Convergence models allow browsers to ask multiple notaries about the legitimacy of SSL certificates [27][38]. The Friends Troubleshooting Network allows a machine to look for mis-configurations based on the configuration policies of machines trusted by users. In our paradigm, we advocate the use of social influence by more directly involving users in doing the overseeing activities [34].

Community oversight allows for social influence – the fact that people are likely to behave like those around them. Social influence pressure can be both explicit, through direct requests for compliance, and indirect, through conformance of behavior amongst group members [9]. If people know there is oversight over their actions, they are likely to change their behavior to conform to the expectations of those around them. Limited social influence mechanisms have been proposed within the domain of information security. For example, DiGioia and Dourish proposed the use of social navigation to aid users in making security decisions, by providing feedback about the decisions others have made [13]. For example, users trying to determine whether to share a particular folder are provided with a visualization about the percentage of other people who also shared the folder, helping the user make a more informed decision. This approach has been further examined by Goecks, Besmer, and Page [1][17][28] for cookie, firewall, and privacy settings. While social navigation mechanisms do allow users to determine what the social norms are within a community, there is no community oversight or feedback, reducing the motivation and ability of users to establish and monitor good security practices. We aim to close that feedback loop, with people more actively watching over each other.

The most related research effort is Chetty et al. [7] where families were able to visualize and monitor home network bandwidth and network usage, as well as customize the identifiers of computers on that network. Users were able to understand and regulate the behavior of family members to identify general network issues and prioritize the bandwidth given to different family members at any time. Users also noted a security benefit in that they were able to detect when unknown computers were using their network, since they stood out from the custom identifiers the family members had chosen. This example demonstrates that by providing users with salient views of the underlying network infrastructure, they were able to utilize this information for monitoring and controlling usage. We similarly argue that by providing communities of users with views of security-related information they will be able to regulate their behaviors.

3. SCENARIOS OF USE

We believe our approach can be applicable to a wide range of security systems, however the design and use of such mechanisms will differ depending on the context of the user and community, and the types of security information involved. Thus, we now present several different scenarios, highlighting possible uses in large organizations, general social networks, and small family groups, before further discussing the general scope and issues of our proposed approach in the next section.

3.1 Organizational

Organizations have defined membership, formal security policies and often a variety of specific people and processes to manage information security. Thus, in this context, we explore a scenario where decisions are by nature distributed and difficult to automatically enforce. Data leakage of all sorts, both unintentional and malicious, is a concern for many organizations, as can be seen by the rise of the data leakage prevention market over the last five years. The concern covers various sorts of information, including business plans, customer information, future strategy and products, and trade secrets. It also covers all sorts of vectors for leakage, starting with email and web interactions such as social

media, through printers and mobile devices. Unintentional sharing may be caused by user mistakes, poor transparency in the user sharing experience ("who is in that group?"), or defects in code. There are a number of ways that community oversight might help with data leakage.

The community that has access to an artifact, such as a file, is a natural one to provide oversight and notice mistaken data leakage. This can be viewed as a "task based community": a community that arises or is created for a specific goal, activity, or task. While users already have interfaces for determining file access, instead of hiding that information in a place the user needs to explicitly look for it, more salient notifications could be designed. A naïve method of oversight could notify everyone with access to a file when someone new has been given access. Many people would find such notifications uninteresting, and potentially overwhelming for very active files. A better way to enable oversight is to embed awareness of who a file is shared with in the experience of interacting with it. This approach can embed oversight into the user's natural tasks, instead of relying on other reasons for a user to check out who has access to a file.

To facilitate such ambient awareness, visualizations can be developed to provide aggregated information, allowing users to notice unexpected patterns or unexpectedly large audiences. Another approach could be similar to Lieberman and Miller's work on displaying faces in email to prevent misdirected email [24]. They show how a set amount of screen space can give useful feedback with photos of recipients. In the case of noticing data leakage, this security-related awareness feedback might always highlight the person(s) a file was most recently shared with, or who most recently downloaded it. Knowing that this sort of feedback is occurring can also spur users who might themselves think they should not have access to a file (either because of a mistake or a coding defect) to take action.

A different form of highlighting might emphasize organizational distance. For example, if every file has an owner, and it is possible to share files both internally and externally to the owner's organization (for example, in cloud based file sharing), users external to the owner's organization could be highlighted. This aligns with the existing data leakage market, which concentrates on leakage across company boundaries. More information can further hone the potential utility of the highlighted information. It could emphasize users who are external to both the owner's and the viewer's organizations. It could use organizational information instead of user faces, and highlight when users in a new organization have access to a file. It could use organizational chart information or user geography information to highlight when a file is shared distantly. It could take into account all the existing users who can access a file when a new user is added, to compute some notion of distance from all users who could previously see the file.

Each of these options demonstrates different methods of compiling and presenting sharing information in ways that will call attention to unusual situations. Beyond noticing that some surprising and perhaps mistaken sharing has occurred, the effectiveness of the technique relies on some sort of follow up action that might inform or remedy the situation. In the case of file sharing, an easy way is to send a notice to the file owner. To expand on the above example, clicking on one of the photos next to testplan.txt might take identity information from that photo and generate an email to the file owner with a subject such as "I noticed Sam Katz has access to testplan.txt". If multiple people are allowed a share a file, tracking of who shared it, and including that person in the notification, may make it more effective. The potential for reporting to other sorts of authority exists, but is likely to diminish both the feedback and the utility of it in a number of ways due to the distributed nature of file permission decisions. Notification features might be augmented by the ability of anyone with access to a file to revoke or suspend access by anyone else (with the subsequent ability to re-enable it). Such powers would only be appropriate in ad hoc sharing scenarios.

3.2 Family

The family (local or extended) can provide a natural community for benign and supportive oversight of more personal security sensitive activities. Many families already place computers in family spaces, set limits on Internet usage, and utilize commercial software to monitor and regulate children. We believe we can extend this oversight into security-related activities, including visiting and sharing information with new websites, ecommerce, and protection of personal machines. A family could also be considered a task based community when the focus is protecting a shared resource such as computers or money. Family members can notify others if they are doing something considered to be unsafe, or can guide others to safer options when there are alternatives or choices.

In the simplest scenario, the family could build up a history of web sites that family members have visited, including information on who has done so, and what additional security sensitive actions they have taken (e.g. registration, personal information input, purchases). This history is then available to family members either explicitly, or as a form of contextual feedback. Different ways to show the information allow for different forms of oversight and utility. Only "new" types of actions might be displayed (a site no one in the family has visited before, the first person in the family to register at or purchase from a site). For oversight in context, references to sites might include either comfort or danger signals based on the usage or feedback of other family members.

This information could allow family members to see where others have been (or acted) in returned search results, or links from other pages [15]. They could then drill down on that information if they felt they needed to contact a family member about potential problems. Oversight could be provided by all members of the (extended) family, or by selected members. In many families, an authority and protector in the context of computers and the web emerges. Individual family members may select different configurations of who watches over them.

For family members desiring more proactive oversight, they could designate one or more "experts" in their family to vet any new site visits or actions before they are allowed to take them (perhaps with the ability to personally override the restriction as well). More and more computer savvy people are becoming such "guides" for their older generation; embedding this oversight into systems can be freeing to those who are particularly cautious. More detailed oversight might include a log of all personal information submitted to sites or, conversely, a registry of information to watch (such as email names and credit card information), with alerts or annotations on who has provided them where. Similar oversight could be applied to other items, such as errors or warnings from applications such as firewalls and email, or email content that a family member finds confusing or questionable.

The privacy tradeoff for these scenarios is particularly interesting, since exposing information to (potentially selected) family members may be a way to achieve better privacy in relationship to

external entities. For example, consider a family member, Charlie, who sometimes accidentally turns off the virus checker. Another family member, Johnny, watches for this, and provides reminders to keep the virus checker turned on. Thus, even though Charlie's actions are more visible and less private, everyone in the family may feel more secure as a result. We discuss privacy more generally in a later section.

3.3 Social Network

Many of us spend significant time interacting online through social media, providing articulated social networks of people to build upon for this approach. A natural scenario would be to use oversight by one's social network over the security and privacy settings of the social network site itself. The people in our social networks are the recipients of the information we share, and could provide feedback when they think someone may be oversharing sensitive information. Users could be provided views of their social network's settings compared against their own settings, showing where users are more or less open than those they interact with. Note that there is not one "right" setting here, users may desire more or less open settings for a variety of social reasons – they may wish to self-promote more than their friends, or decide they are more private than many of their friends. But currently, it is difficult to judge how one's settings are in relation to the social norms, and to spot others who perhaps need assistance in following better practices. For example, many users may be unaware of the danger of sharing information such as birthdays and phone numbers that can be used for identity attacks. Users could both promote their own settings amongst their social circles and decide to follow other people whose behaviors they wish to emulate. Bonneau et al. suggested a related technique, where users could post privacy settings "profiles" that others could follow, reducing the burden of configuration [5]. Our approach would make such sharing more implicit and social, allowing users to determine and effect how their settings align within their social circles.

Another possible application is to utilize such social network and sharing features of a site to share security information unrelated to the social media site. For example, an application could allow people to post and share which types of security-related applications they use on their personal computers and devices. This could help users promote various security products and allow users to watch out for poor products (such as fake anti-virus programs) and for those who choose to use too few products. A security consideration of such sharing is that it could provide a pool of information on users (and systems) that are underprotected and hence better targets for certain classes of attacks.

In both of these scenarios, the privacy of the settings or products used may be an issue depending on the size and makeup of one's social network, and on the inferences people make about each other based on the added knowledge of security and privacy behaviors. Currently, I may realize that I am not seeing any photos from John, but do not know if its because John just doesn't post photos at all, or does not choose to share photos with me. Thus, we must take care in sharing settings that indicate the strength of relationships between people, and either aggregate or generalize such settings or focus on information that will not indicate any additional information about the social relationship.

4. SCOPE AND CHALLENGES

We look at the issues with this paradigm in two parts: the people involved, and the policies and procedures that need to be in place. Within each we provide a framework of aspects that can be altered within the overall system. The community is defined by its structure, size, cohesiveness, the relationships between the members, who the authorities are, and how watchers are identified. Aspects of policy beyond the actual content that are of specific concern include the incentives built into it, the technology that supports it, and privacy implications. This provides a framework for further research on this paradigm. The aspects of Community Structure can be mixed and matched with those of Policies and Practices, to discover where oversight might be the most effective.

4.1 Community Structure

How the community is defined and **structured** can have a large impact on many of the aspects of what oversight for security might accomplish. A contractually specified organization, such as a business or not-for-profit, can have clear boundaries, integration with a central authority, pre-defined roles and responsibilities, and specific work groups and reporting hierarchies, all of which can be leveraged in various ways. At the other end of the spectrum are self-organized communities, based on personal ties, interests, or other sorts of social networks. People adopt and participate in these communities as they see value, which means that articulating the value of a form of security oversight is a central challenge to their use in this context. The communities themselves may evolve over time, as may the value of security oversight. An additional type of community would be one defined by more extrinsic ties, such as a household, family, or a particular geography (neighborhood, town, campus). Sharing an IP address or having nearby IP addresses can also be thought of as extrinsic tie communities, since bandwidth and blacklisting may affect all members. These different types of communities will have different goals and social dynamics, making them appropriate for different types of security oversight.

An important variable in these different communities is the **size** of the community. Smaller communities may lead to more quality or (warranted) trust; larger to better coverage or significance. In an entirely homogeneous community where everyone watches everyone else, scale would pose problems. The number of relationships would be (n*n-1)/2, scaling geometrically in the worst case, posing a challenge for determining understandable outcomes and feedback, not to mention the actual system performance implications.

Size may also determine **relationships** among the members of the community, including whether they know each other, know of each other, or have any other relationships or affinities. People in smaller communities are likely to have stronger social ties, which exert greater social influence. The **cohesiveness** of a community will also determine the level of engagement of the members. The more people identify themselves with the community and its social norms, the more likely they will be to engage in security oversight and comply with those norms to gain group acceptance. For example, there is anecdotal evidence that a strong sense of community and ownership in a SecondLife sims can result in effective oversight of shared resources such as server load. However, while we hope that the existing social norms in the community can influence the direction of security oversight to a higher chance of success (or stronger result), it could also impede potential benefits if community-norms are misguided or view good security practices negatively. In addition, variations within the community are always likely to exist, with some members having less trust in the community, or an explicit negative reaction to any forms of authority. Social trust is a core aspect of

this paradigm; the acceptance of the risk that comes with social trust varies by individual and by community.

The structure (hierarchical, flat) and roles of people within a community will also impact the type and outcome of oversight. There may be substructures within the community if it is big enough, with different levels of engagement or trust. Such substructures might be used to answer the traditional question, "Who will watch the watchers?", much like various forms of broader community oversight have been proposed to watch those who monitor others.

Another important aspect of community structure is whether there is an explicit or implicit **authority** in the community, to determine the security goals and to request compliance with those goals. It could be the user at the "center" of such a network, the owners of the technology platform, or someone else. Within a formal organization, security authorities are often defined in various job roles. Employees or members of the organization will likely accept that authority for job-related security tasks, such as preventing data leakage of corporate files. However, they may not acknowledge the same people as having authority over more personal security tasks. In less structured communities, individuals may assert authority and groups may grant authority to community members based on various social factors. For example, in one UK community researchers uncovered that 40-something year old grandmothers serve as authority figures for their grandchildren on computer privacy issues such as Facebook disclosure behavior [11].

Finally, how the community determines who functions as "**watchers**" is critical. Watchers could be determined by an authority or community or self-selected based upon personal traits or knowledge. What they need to know, and how that knowledge is acquired (training, experience), particularly around security, is a key concern and will differ depending on the security tasks and communities involved. Their relationship to both the community and to the watchees also needs to be considered. They may take on some formal or explicit role within the community and that role may determine who they are responsible to, or who they report to. In our Neighborhood Watch analogy, some organizations have block captains responsible to a block organizer. Watchers may be formally recognized as such, or chosen implicitly based upon who community members decide to trust and listen to. If watchers have no formal role, then there need to be other mechanisms for their efforts to be recognized and valued by the community. If watchers are chosen by the community, the social dynamics of the group needs careful consideration. To what extent would it be like an election or popularity contest? How would the traits the users would select based on (perhaps personal trust, interpersonal similarity, over desire to do the job) compare with the quality that would make an effective or useful overseer in some context and for some policy? A related dimension is if knowing who is overseeing you changes your behavior "around" your overseers. As certainty of detection has been shown to be an incentive for security compliance [19], in some contexts there might not be enough watchers to sufficiently detect issues to be effective or add value.

The potential for social engineering (as well as other attacks) by both the watchers and the watched is an important consideration for future work in this area. Misuse cases [32] is one method for looking for them in any particular instantiation of this pattern. In our organizational scenarios, and in classic intrusion detection fashion, a user who actively plans to leak information can create a plausible and benign norm for the type of person she wishes to leak it to. Family "experts" could easily seed their shared history with sites they know to be unsafe, as a prank or with more directed malicious intent. No social system can guarantee all members behave, yet any effective community develops varying ways of dealing with such threats, and we believe so can communities of oversight.

The least structured and most dynamic communities, online social networks, have the most varied potential for social engineering. They might also have the most potential for emotionally reactive abuses of power, such as "witch hunts." Social network software is continually evolving and changing what information can be seen by whom, and explicitly builds in notions around transitivity (e.g. friends of friends). Social networks can easily "bleed into" more structured communities, such as the family. In all cases, there is the potential for information to flow through overseers to unexpected and unknown others. In these cases, watchees (or authorities) who are trusting overseers are then also trusting other people who the overseers trust. Research in transitive trust [21] can provide some initial models for considering the impact and implications of these kinds of flows. There is the potential for transitive trust to work positively; to provide excellent and trustworthy overseers removed from the watchee. Vetting and choice by the watchee, or other forms of expert designation, can provide that advantage. Research on the power of communities will be critical for setting up the right watcher dynamics, as well as consideration of privacy (see below).

4.2 Policies and Process

The potential security goals and benefits of oversight will depend not only on community structure, but what can be achieved or enabled by the types of policies and incentives that are supported by the tools and technologies.

4.2.1 Policy

An important area for consideration is what policies oversight is applied to. The policies need to be something that overseers can understand and relate to, or that oversight that is possible relates to. The question of who sets the rules for what is good practice or policy can get intimately tied up with the effectiveness of the oversight. In an organization, some policy is specifically set, communicated and documented by the parts of the organization responsible for security (e.g. management, CIO's office, IT, HR). Other forms of good security practice are not clearly articulated, whether the community is a structured organization, a family, or a social network. And some areas of security best practice are hotly contested on an individual or cultural basis. Popular advice may be stubbornly wrong. Choice of policy will need to align with the community, the overseen and the overseers. Community members and watchers will also need to understand the need for the policy and its benefits to care enough to oversee and be overseen. For informal policies, questions of security and privacy expertise and education loom large. Work in the healthcare arena on empowering community leaders and "training-the-trainer" may provide some basis for education and valuing of appropriate security expertise [12].

For example, in our previous file sharing example, an organization may be able to articulate specific policies regarding sharing information externally. Users would hopefully be able to both detect violations of such policies, as well as understand necessary exceptions. Within a social network community an example policy might be to not share birthdays and contact information with anyone other than close friends. Yet such

policies will not be clearly articulated, may vary and conflict across the community, and change over time.

One aspect of policy selection is the potential for false positives and negatives. This can be a consideration for many security mechanisms, but seems particularly of interest in a system where people are providing the inputs and decisions. There will need to be ways to separate good feedback from bad feedback, and depending on the policy and overseers, there may be many more false positives and negatives than true [1]. Reputation systems, redundancy, and analytics may all play a role in sorting through false hits [25] and allowing users to judge the decisions and feedback from the community. Another noteworthy devolution of policy is its reduction to compliance rules, which can be interpreted and acted on in ways that can be ineffective or even counterproductive.

4.2.2 Technology

The technology in use will certainly influence oversight, and security mechanisms may be changed to enable desirable oversight. A number of technologies can be used to define a community, including existing social network sites such as Facebook, LinkedIn, GooglePlus, and Twitter. A number of mechanisms used in organizations and elsewhere can also define a community, including email groups, organizational directories, and groups used to set up sharing and access to business applications. Technology is also required for communication between community members, which may include direct discussion as well as views of security-related information about others.

An important aspect of the use of technology for oversight is what information can be seen by the overseers. The key to this approach is providing salient views of the underlying security-related information. This means gathering, aggregating, and presenting such information in useful, non-intrusive ways. A good deal of information is potentially available in existing frameworks. For example, information about sharing is available (to some extent) as the friends list in Facebook, connections in LinkedIn, and in display of access control information in most information sharing applications. Information around other security aspects, such as authentication and configuration, is less available. To what extent can existing security infrastructures be leveraged? Is it simply a matter of making things more available and transparent? Or is there a need to do additional monitoring and reporting to people? Several dimensions may drive the need for data tuned to the oversight task, including privacy concerns, and mapping overseer understanding of security to the data at hand. Analytics may be an approach that can provide a shared understanding of the broader security behaviors of the community, and individual security decisions may be compared with those.

Information about the interactions among community members may also highlight useful patterns and aberrations. Community membership through social network connections, familial relationships, organizational membership, org charts, and group definitions, as well as overlap between community members can bring transparency to the actions between community members in addition to their interaction with information.

Another aspect of the technology is what actions an overseer takes. We can consider how technology might make oversight easier, not just in terms of transparency of actions, but also follow up. Just as "like" and "poke" are one click actions, a "look out" action might signal overseer concern quickly and simply directly to the person of concern. If an indication of a policy transgression, such as oversharing, is as light weight to mention to someone as a "like" button, and is private to the user being flagged, how might that encourage and enhance oversight? We believe that solutions could be explored that build such security-related oversight features into existing social systems, such as social networks sites, and that add social and community features to existing security systems.

4.2.3 Incentives

Herath and Rao have investigated incentives that influence adherence to security policies within organizations [19]. They demonstrate that subjective norms, peer behaviors, intrinsic motivation, perceived effectiveness, and certainty of detection have positive effects while severity of punishment has a negative effect. While our approach provides a mechanism to enable such incentives for security behaviors, the mechanism itself will likely need application-specific incentives. Whether or not someone is a "watcher", they must still be willing to join the community, find value in the feedback they receive from others, and be willing to modify their behavior based on that feedback. While a formal incentive structure is possible, a social system is likely instead to depend upon social benefits and those will have to be designed into and supported within the system.

If oversight imposes additional work or attention from watchers, why would they do it and what benefit would they achieve? What activities and threats will they care about? Mackay [26] found that in groups working with customizable software, a "translator" emerged from the community, and that translator was by and large not the person in authority or with the most obvious technical skills. Similarly structured studies may highlight what sort of personal traits may determine effective or engaged overseers. A desire for a sense of connection, an interest in the business of others, or a sense of responsibility or authority are all potential candidates. Research in online communities has also uncovered a number of important design aspects that can promote similar kinds of community participation, including that people contribute more when they see their contribution as unique and important, when the group benefits are made more salient, and when they are reminded of the multiple and intrinsic benefits of contributing [2]. Linking overseer incentive to work [36] could produce additional insights into potential incentives. Technologies that make security a "club good" may provide built in incentives [14].

4.2.4 Privacy

Oversight, by its very nature, balances privacy with transparency, since something must be shared with others for them to watch. We touch on this issue at the end of the "Family" and "Social Network" scenarios above. We can imagine various controls that might help balance and control privacy concerns in such a scenario, including individual choice of who watches over them (including no one), and the ability to perform a certain class of actions without oversight (such as visiting certain web sites). danah boyd has suggested that public by default, with mechanisms to make things private, is desirable for at least some community and coping strategies [6]. However, data which we consider most private can provide an attractive attack vector which may not be covered by this approach. The extent to which oversight can be concentrated on security-related cues rather than raw informationmay help with this tradeoff. Within a community, one obvious concern is how oversight aligns with or crosses more formal power relationships, such as the management chain, or parent/child relationships. Users need notification of oversight, and also need to somehow be held accountable for any oversight

they are doing. Choosing your overseers may mitigate some privacy issues, but may raise efficacy and power issues. Scenarios where people can watch stuff that they can already see are an attractive starting place, since privacy is less of an issue in them. Our organizational file sharing scenario above is one example, as is the ability to give feedback on the outcomes of privacy settings on social network sites.

There are interesting questions to consider when looking at the benefits and drawbacks of greater transparency. What information is too much, in terms of quantity that can overwhelm or occlude useful data, or that crosses the privacy line of either the overseen or the overseer ("too much information dude"). George Orwell's 1984 commentary on cameras covering the streets of London (and the Simpson's episode parodying that kind of surveillance) all speak to the societally undesirable reactions and changes of total transparency. The potential for social media to make everyone a watcher, and to make the watchers watch each other, recalls small community living, where everybody knows your name, and you know theirs. Understanding and mitigating the potential privacy concerns will be necessary for a successful system.

5. CONCLUSIONS AND FUTURE WORK

One of the major challenges in security is influencing people to actually follow security policies and behaviors. While research has examined how organizations can implement policies to bring about these behaviors, few other communities have such formal and structured policies. Improving the usability of security technologies can certainly help, yet users often still need motivation and assistance for security-related work and decisions. In this paper we propose community oversight as a promising and interesting approach that could be widely applied to address these issues. The fact that so many examples of similar group oversight exist for physical-world security and safety make this approach an intriguing method worthy of further discussion, exploration, and research. We have demonstrated that this method can be applied to a number of communities and security decisions with a variety of aspects that can be varied across solutions for different impacts.

The integration of social and community mechanisms into security technologies presents a variety of issues that should be studied further as concrete systems are designed and evaluated. One key issue will be in the interface and interaction design for several areas: getting people to notice anomalous behaviors, providing lightweight visualizations of security information, and easy mechanisms for communication and action. For example, taking Patrick et al.'s [29] work on designing for trust through consistency and communication as a baseline, and turning it on its head, can yield some indicators of what kinds of anomalies people might notice and think worthy of distrust. Chetty et al. also point to what users can notice through oversight (e.g. items that have not been personalized or otherwise marked as "normal") [7]. Wolgater's Communication-Human Information Processing Model (C-HIP) [39] provides a structure for research on response to computer security warnings [16] and the same structure can ground future research on what overseers would notice, understand, and act on.

The effectiveness of community-oversight mechanisms will depend on many aspects of the design, including the issues of community structure and policies we have discussed, as well as the actual decisions and collective behaviors of the community members. While this may seem at odds with the more automated traditional security methods which can be more straightforward to model to predict performance, the growing focus on analytics and modeling of social systems can provide insight and methods for implementation and evaluation of a community-oriented security system. Emerging research in social network interaction on Twitter shows some of the attributes of social networks that can be analyzed and modeled [40]. One related area of study is the question of diversity vs. monoculture in security mechanisms. If a diversity of security approaches and policies are used by different communities, are they resistant to some forms of attack? As with any security approach, how attack resistant it is will be proved in part by a concerted effort to attack it. Related to that would be work on confounds and bleed through for membership in multiple communities.

There are also research efforts at incentive models for community and group oriented systems that can be extended by looking at incentives for all community members to participate. For example, gamification is a technique to make a non game activity more engaging to users. We can imagine using a social or organization community as the set of participants, and turning password strength into a game. Community averages might become the baseline (with some degree of freedom, presuming they are above a certain threshold), and kudos might go to the members with strongest passwords. We posit this approach might do well in anonymous communities. Marketing studies indicate the strongest influence on people's behavior can be descriptive normative information (what the neighbors do) [10], which may tap into the same competitive impulses. However, the potential for competitions to become a "race to the bottom" instead should be a consideration for structure and feedback.

Security technologies already operate in a complex socio-technical system of people and organizations. We believe that by incorporating more of this social context, we can augment security mechanisms and influence users towards stronger security behaviors. We plan to continue to examine this new paradigm, researching the specific issues we have raised as we design and evaluate concrete technologies.

6. ACKNOWLEDGEMENTS

We would like to thank Steve Greenwald for his early review and insights. The participants at NSPW 2012 who provided useful input to the final version of this paper comprise the bulk of the attendees. Thank you all for your thoughtful feedback.

7. REFERENCES

[1] Augustin, Eriq, Cailin Cushing, Alex Dekhtyar, Kevin Mcentee, Kimberly Paterson and Matt Tognett. Outage Detection via Real-time Social Stream Analysis: Leveraging the Power of On-line Complaints. To appear in WWW2012.

[2] Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., & Kraut, R. E. 2004. Using social psychology to motivate contributions to online communities. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 04)*.

[3] Besmer, Andrew, Jason Watson, and Heather Richter Lipford. The Impact of Social Navigation on Privacy Policy Configuration. In *Proceedings of the 2010 Symposium on Usable Privacy and Security* (SOUPS 10).

[4] Beautement, Adam, M. Angela Sasse, and Mike Wonham. 2008. The Compliance Budget: Managing Security Behavior in Organisations. *New Security Paradigms Workshop* (NSPW '08).

[5] Bonneau, Joseph, Jonathan Anderson, and Luke Church. 2009. Privacy Suites: Shared Privacy for Social Networks.

Poster at *The Symposium on Usable Privacy and Security* (SOUPS '09).

[6] boyd, danah. CFP11 Day1 Keynote danah boyd. http://www.youtube.com/watch?v=W6BNCZlsIuQ.

[7] Chetty, Marshini, David Haslem, Andrew Baird, Ugochi Ofoha, Bethany Sumner, and Rebecca Grinter. 2011. Why is my internet slow?: making network speeds visible. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). pp 1889-1898.

[8] Chia, Pern Hui and John Chuang. 2012. Community-based web security: complementary roles of the serious and casual contributors. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (CSCW '12). pp 1023-1032.

[9] Cialdini, Robert B. and Noah J. Goldstein. 2004. Social Influence: Compliance and Conformance. *Annual Reviews Psychology* 55: 591-621.

[10] Cialdini, Robert and Wesley Schultz. 2004. Understanding and Motivating Energy Conservation via Social Norms. Project Report.

[11] Coles-Kemp, Lizzie. 2012. Privacy Awareness: An Inclusive Approach. Oxford Internet Institute webcast, August 15, 2012. Available at: http://webcast.oii.ox.ac.uk/ ?view=Webcast&ID=20120815_448.

[12] Cuaresma, Charlene, Diane Mitschke and Hali Robinett, Capacity Building for Cancer Awareness in Hawaii's Foreign-born Filipino Communities. In *Developing Human Resources in the Pacific*. Vol 14, No. 1, 2007.

[13] DiGioia, Paul and Paul Dourish. 2005. Social navigation as a model for usable security. In *Proceedings of the 2005 symposium on Usable privacy and security* (SOUPS '05). pp 101-108.

[14] Dong, Zheng, Vaibhav Garg, L. Jean Camp and Apu Kapadia. Pools, Clubs and Security: Designing for a Party Not a Person. In *Proceedings of New Security Paradigms 2012*. To appear.

[15] Egelman, Serge, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. 2009. Timing is everything?: the effects of timing and placement of online privacy indicators. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09). Pp 319-328.

[16] Egelman, Serge, Lorrie Faith Cranor, and Jason Hong. 2008. You've been warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proceedings of the 26th international conference on Human factors in computing systems* (CHI '08).

[17] Goecks, Jeremy, W. Keith Edwards, and Elizabeth D. Mynatt. 2009. Challenges in supporting end-user privacy and security management with social navigation. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (SOUPS '09).

[18] Guenther, Melissa. 2004. Security/Privacy Compliance: Culture Change, EDPACS, 31:12, pp 19-24.

[19] Herath, Tejaswini, and H.R. Rao, 2009. Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness, *Decision Support Systems*, 47(2), May 2009, pp 154-165.

[20] Holloway, Katy, Trevor Bennett, and David P. Farrington. 2008. Crime Prevention Research Review No. 3: Does Neighborhood Watch Reduce Crime? Washington, D.C.: U.S. Department of Justice Office of Community Oriented Policing Services.

[21] Josang, A., E. Gray and M. Kinateder. 2003. Analysing Topologies of Transitive Trust. In *Proceedings of the Workshop on Formal Aspects of Security and Trust (FAST).*

[22] Kandogan, Eser, and Eben M. Haber. 2005. Security Admnistration Tools and Practices. *Security and Usability*, Cranor and Garfinkel eds, O'Reilly Media Inc.

[23] Kraut, Robert E., Ronald E. Rice, Colleen Cool and Robert S. Fish. 1998. Varieties of Social Influence: The Role of Utility and Norms in the Success of a New Communication Medium. *Organization Science* , Vol. 9, No. 4 (Jul. - Aug., 1998), pp. 437-453.

[24] Lieberman, Eric and Robert C. Miller. 2007. Facemail: showing faces of recipients to prevent misdirected email. In *Proceedings of the 3rd symposium on Usable privacy and security* (SOUPS '07). Pp 122-131.

[25] Liu, Gang, Guang Xiang, Bryan A. Pendelton, Jason I. Hong, Wenyin Liu. 2011. Smartening the Crowds: Computational Techniques for Improving Human Verification to Fight Phishing Scams. In *Proceedings of the 2011 Symposium on Usable Privacy and Security* (SOUPS 11).

[26] Mackay, Wendy E.. 1990. Patterns of Sharing Customizable Software. In *Proceedings of the 1990 ACM conference on Computer-supported Cooperative Work* (CSCW '90).

[27] Marlinspike, M. 2001. SSL and the future of authenticity - moving beyond certificate authorities. Video recording: Blackhat USA 2011, Aug. 2011.

[28] Patil, Sameer, Xinru Page, and Alfred Kobsa. 2011. With a little help from my friends: can social navigation inform interpersonal privacy preferences? In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (CSCW '11). pp 391-394.

[29] Patrick, Andrew S., Pamela Briggs and Stephen Marsh.,2005..Designing Systems That People Will Trust. *Security and Usability*, Cranor and Garfinkel eds, O'Reilly Media Inc.

[30] Pieters, Wolter and Lizzie Coles-Kemp. 2011. Reducing Normative Conflicts in Information Security. In *Proceedings of New Security Paradigms Workshop 2011*. pp 11-24.

[31] Postmes, Tom, Russell Spears, Khaled Sakhel, and Daphne de Groot. 2001. Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behaviors. *Personal and Social Psychology Bulletin*. October 2001. 27: 1243-1254.

[32] Sindre, G and A.L. Opdahl. 2000. Eliciting Security Requirements by Misuse Cases. In *Proceedings of 37th International Conference on Technology of Object-Oriented Languages and Systems*. pp 120-131.

[33] Vroom, Cheryl, and Rossouw von Solms. 2004. Towards information security behavioral compliance. *Computers & Security*, 23, pp 191-198.

[34] Wang, Helen, Yih-Chun Hu, Chun Yuan, Zheng Zhang, Yi-Min Wang. 2005. Friends Troubleshooting Network: Towards Privacy-Preserving, Automatic Troubleshooting. In

Peer-to-Peer Systems III, Lecture Notes in Computer Science, Vol. 3279/2005, pp 184-194, Springer Berlin/Heidelberg.

[35] Wang, Yang, Gregory Norcie, Saranga Komanduri, Pedro Giovanni Leon, Lorrie Faith Cranor and Alessandro Acquisti. "I regretted the minute I pressed share": A Qualitative Study of Regrets on Facebook. In *Proceedings of the 2011 Symposium on Usable Privacy and Security* (SOUPS 11).

[36] Wash, Rick and Emilee Rader. 2011. Influencing Mental Models of Security: A Research Agenda. In *Proceedings of New Security Paradigms Workshop 2011*. pp 57-66.

[37] Web of Trust add-on. http://www.mywot.com/.

[38] Wendlandt, D., D. Andersen, and A. Perrig. 2008. Perspectives: improving SSH-style host authentication with multi-path probing. In *Proceedings of the USENIX 2008 Annual Technical Conference*, pp 321-334.

[39] Wogalter, M. S. Communicatin-Human Information Processing (C-HIP) Model. 2006. In *Handbook of Warnings*, M. S. Wogalter, Ed. Lawrence Erlbaum Associates, pp. 51-61.

[40] Wu, Shao mei, Jake M. Hofman, Winter A. Mason and Duncan J. Watts. 2011. Who Says What to Whom on Twitter. In *Proceedings of WWW2011*. pp 705-714.