The Ethics of Social Honeypots

David Dittrich
Applied Physics Laboratory, University of Washington *

Abstract
It is the nature of research that knowledge gained from one study is built upon in succeeding studies. The lessons learned – be they good or bad – help future researchers to find their path. At the same time, publication of research results is difficult when publishing imposes page limits that require leaving things out. Such restrictions in what can/can’t be included may leave unstated very important issues regarding risks to avoid, harms that were mitigated, or responsible conduct of research procedures that those who are new to a field would benefit from knowing.

This paper considers some of the ethical issues surrounding the study of malicious activity in social networks, specifically using a technique known as social honeypots combined with deception. This is a potentially touchy area of study that is common to social and behavioral research that is well understood to fall within the boundaries of human subjects research that is regulated in the United States and reviewed by Institutional Review Boards, but is not well understood by computer security researchers or those in the private sector.

1 Introduction
Users of the internet are constantly under attack by malicious actors who are intent on stealing login credentials, bank account numbers, credit card numbers, etc. The tactic that is most blatant is known broadly as social engineering, deceiving users by pretending to be someone who the victim trusts and asking for users to give up sensitive information. The victims are often approached by email messages, direct messages, tweets and status updates. This is known as phishing [13]. A classic phishing messages goes something like this: “This is the Security Department. We have noticed abnormal activity with your account. Please validate your identity immediately by giving us your name, address, Social Security Number, and date of birth. Failure to validate your account in the next 48 hours will result in IMMEDIATE TERMINATION of your account! Thank you for your cooperation.”

A more subtle means of targeting victims is to simply send what looks like a link to a video or post about a current topic. After an earthquake or tsunami, it may be a video of someone in an amazing rescue, or horrific accident. The link sends the victim to a compromised web server that has a browser exploit kit [2] on it that identifies one of dozens of vulnerabilities on the victim’s computer and forces it to load malicious software. The victim may have no idea their computer was just compromised, yet the next time they log in to their bank, the bad guys get to know about it.

Social networks, such as Twitter and Facebook, are now used by millions of people around the globe and are becoming influential in social movements, such as pre-election “get out the vote” campaigns and the Arab Spring protests. They are prime targets for phishing attack, in this case it is URLs sent in tweets, status updates, and direct messages from already infected accounts, that can cause the problems. A program named Koobface [22, 19] is such a threat to Facebook users.

1.1 Privacy of Personal Data
Social networks serve many purposes, primarily focused on both public and private communication. They link friends and family into groups, or provide a simple means of speaking out on a topic in a way that makes it easy for others to follow. They are growing in popularity as a means of keeping in touch with others around the globe and are used by some people on a nearly constant basis via smart phones and other personal electronic devices.

Users of social networks who restrict access to their posts to only “friends” have made a decision to limit what posts are made public and thus have an expectation that some communications and personal data they store in the social network platform are private. Privacy of personal data is a serious subject and the Federal Trade Commission (FTC) in the United States has brought a number of law suits against corporations for privacy rights violations. In one case, Google agreed to a record $23M settlement for violating its user’s privacy rights [21] and Facebook settled an agreement over failing to honor privacy promises in its user agreement [9]. A third law suit filed against a company named Compete

*The author serves as Chief Legal and Ethics Officer of The Honeynet Project and is an affiliated scientist member of an IRB Committee at the University of Washington.
charged that consumer tracking software surreptitiously collected credit card and Social Security numbers from consumers who downloaded their software, as well as failing to properly protect the sensitive data it had collected [20]. Individuals whose private data is collected by corporations without permission, or used in ways that go beyond agreed upon terms of service, anger people because their autonomy in deciding how to protect their personal data, or to opt out of allowing their sensitive data to be used in ways they do not approve of, has been taken away from them.

The classic honeypots – computers that were placed on the internet for no other reason than to allow someone to compromise it in order to learn how attackers operate and to counter those attacks – by definition had no users, and thus no personal communications or personally identifiable information (PII). There was therefore no need to get anyone’s permission to collect, use, or disclose any data collected. The exact opposite is true of social honeypots, which may be filled with potentially private communications and PII. This distinction is sometime lost on researchers who wish to enter into social honeypot research, focusing on the honeypot and not the context in which it exists and how it is used.

1.2 Use of Deception in Research

On top of the issue of the collection, use and disclosure of PII, however, is that of deception.

Deception in research is often necessary in order to perform social and behavioral studies that involve things like actions based on conscious or subconscious decision making. If you were to inform a research subject that you were about to provide them with false information, this knowledge could very well change their choice and ruin the scientific validity of any observations.

Deception sometimes is used in computer security research, such as in the study of social engineering. It is unlikely a user will be tricked into giving their password to someone pretending to be an authority figure if they must first sign a consent form that explains, “In this study, we will send you fake email messages in which we pretend to be your bank and ask you to give us your account number and password in order to learn how easy it is to trick you.”

Researchers within the Honeynet Project have long discussed applying the concept of honeypots into the social networking realm, using social honeypots to detect attacks such as KooVface. Other researchers in academia have similarly picked up on this idea and are pursing the concept [14, 23].

In this paper, the ethical issues and challenges of researching threats to social networks using social honeypots are examined in hopes that future research will benefit from understanding these issues and how to address them.

2 Social Network Defense Research Case Studies

Social networks are designed both to link users together into a fabric of friend and/or follower relationships, as well as to facilitate communication between these friend/followers as part of daily life. As such, they are clearly a target for malicious actors who wish to exploit weaknesses in the social network to send malicious messages designed to cause harm, but to come in the guise of normal friendly communication. This is known as social network spam. Lee et al [14] and Grier et al[11] both address the issue of social network spam detection using the honeypot concept of passively monitoring for posted messages and examining them in order to identify malicious communications.

Both the Lee and Grier studies involve collecting data from only public (i.e., non-private) communications. As such, these are not private communications that carry any expectation of privacy that could subject them to IRB review. This fact is not clearly mentioned in either paper, however, and the words privacy and ethics do not occur at all (except in titles of cited references.) Anyone reading these papers, who does not already have a strong understanding of the nuance of use of public data in research, might build on this study and extend it to non-public posts without even knowing this raises ethical issues.

A more recent study by Zhu et al [23], cites both Lee and Grier as the basis for their proposed experiment. This study, however, does not focus on just public accounts, but takes the concept of social honeypot and explicitly adds deception in order to build a social network of 10^6 users. This study proposes to use a large number of fake identities and send out invitations to a million users in order to establish a large social network. All messages sent to these fake identities are examined and analyzed to identify malicious spam messages intended to infect users’ computers or steal information from them. From the results of analysis, defensive mechanisms can be deployed to protect the users of the social network.

2.1 Observations

In all three of these papers, the lack of any statements about issues of privacy, ethics, IRB review, etc., all leave silent any guidance for readers. As Miller states, “Dis-
cussing key ethical issues in published articles promotes public moral accountability, just as discussing research methods promotes scientific accountability [17].” The Menlo Report [5] includes a separate principle – Transparency and Accountability – that echoes Miller’s position. Deibert and Crete-Nishihata [3] stress the importance of performing research into widespread criminal activity on the internet, including the “need for explicit research rationales and the use of research warrants.” They similarly suggest a form of pre-research transparency is needed, both to help guide ethical research design, but also to help inform future researchers of how to similarly take the right path. “The best way to ensure that [ethical] standards are met is through careful, clear, and explicit documentation of research methods and justification of the choices that are made along each step of a particular project. Doing so can build up reference points and a knowledge base for future research.” The method they suggest, something called a research warrant, is “written by the principal investigator, [outlining] the nature and justification for all aspects of the research, which [is] then incorporated into the text of the published reports.” One of the main benefits of the research warrant is that this statement is prepared well in advance of the research even starting in case the researcher is called to defend their research activities.

In discussing the 1975 modification to the Declaration of Helsinki calling for unethical research to not be published, Miller states, “The debate over publication of unethical research suggests a false dichotomy. On one hand, if research is unethical, its results should not be published, or published only if accompanied by an editorial condemnation. On the other hand, if research is ethical, then there is no need to discuss ethical considerations. This dichotomy ignores the fact that research might have morally controversial features without necessarily being unethical [17].” The purpose of this paper is not to judge the ethics of these studies. Rather, it is to urge researchers to be open and transparent about issues that may be morally controversial and to, as called for in the Menlo Report, “to enable ICT researchers and oversight entities to appropriately and consistently assess and render ethically defensible research [5].”

One could consider what would happen if these papers were submitted for grading, or submitted for publication to a conference such as the Symposium On Usable Privacy and Security (SOUPS). In the same way that providing an answer to a complicated mathematical proof in a homework assignment or exam requires transparency as to how the answer was derived, such papers could well be marked down for not illuminating the thought process that would have been necessary had these studies been submitted to an IRB for review. Researchers may be well versed in ethics and may have every base covered, or they may have not have even given a single thought to the ethical issues involved. A reader has no way of knowing because of the silence on the topic. If the papers were submitted to SOUPS, the authors would be asked to, “follow the basic principles of ethical research, e.g., beneficence (maximizing the benefits to an individual or to society while minimizing harm to the individual), minimal risk (appropriateness of the risk versus benefit ratio), voluntary consent, respect for privacy, and limited deception. Authors may be asked to include explanation of how ethical principles were followed in their final papers should questions arise during the review process.”

Based on the SOUPS requirement, each set of authors may have at best been asked to explain the ethical issues involved and how they were addressed, may have been asked to use precious page space (or add a link to more detailed documentation) explaining the ethical issues and how they are addressed, or at worst may have had the paper be rejected on ethical grounds. The studies involving public data are much easier to address than a study involving collection of private communications and/or using deception. At minimum, the issue could be raised and clarified with statement as simple as, “This study involves only data made public by those from whom it was collected, which avoids the necessity of review on ethical or privacy grounds.” Asking a researcher to consider and justify the ethical issues in a study at the time of submission of results for publication, however, is itself a risky practice that the research community has a responsibility to address [4]. If any harm does manifest from performing experiments without properly addressing the ethical issues in advance, the users involved will have had suffered harm long before the publication was submitted and rejected.

It is unclear which would be considered a tougher ethical call: To allow deception of 100 research subjects in a consent form they read and signed in acknowledgment of the risks and benefits of the research in advance, or granting a waiver of informed consent and allowing deception of a million users of a social network whose communications are mined for malicious content? Do the number of people involved alone increase the risk, or the burden on the researcher to mitigate potential harms? These are good questions for researchers to ask of an IRB administrator or committee member, preferable prior to engaging in potentially controversial research. (To help promote the discussion of how research such as this could be presented to an IRB, a mock IRB application is presented in
3 Addressing the Ethical Issues

We now look at the ethical issues involved, and explain how they might be addressed within the IRB regime in academic institutions in the United States. Researchers in other countries may have different regulations, or different review bodies, under which they operate. The reader is advised to consider their own situation and act appropriately.

While this discussion focuses on researchers in the United States under the IRB review regime, that doesn’t mean that non-academic researchers (e.g., private individuals, researchers in the private sector, or employees of governmental or non-governmental organizations) can do whatever they want. They are still subject to appropriate use policies (AUPs) of the platforms on which they perform their research studies, and will still be called to answer for actions that are perceived by those who believe their privacy has been violated or have harmed them in some way (as seen in the FTC law suits mentioned in Section 1.1. Nor does it mean that an academic researcher can simply partner with someone in the private sector in order to do their research without having to go to an IRB, unless they are doing their research under the auspices of the service provider and conforming with that service provider’s agreements with its users or customers.

3.1 Stakeholder Analysis

As an initial step in understanding the risks, benefits, and how to appropriately balance the two, a comprehensive stakeholder analysis is helpful. Stakeholders are divided into three categories and have both positive (innocent) and negative (malicious) aspects: Key stakeholders are those with a direct impact on delivering benefits or harms; Primary stakeholders are the principle entities receiving benefits or harms; Secondary stakeholders are intermediaries in the delivery of products or services related to the subject of study, such as service providers or vendors.

3.1.1 Key Stakeholders

Researchers (positive) Researchers gain knowledge from the study of malicious attack mechanisms, and benefit from the development of defensive mechanisms through the development of generalizable knowledge, the the generation of of intellectual property (technology) that can be transitioned into products or services to benefit the general public, and though publicity and further funding resulting from publication of research findings. They are also potentially harmed through negative publicity, having to respond to complaints from primary or secondary stakeholders who are harmed, and even by attacks from negatively inclined key stakeholders [10]. In some cases, researchers’ actions may affect the results of other researchers’ experiments [6] as these studies are being performed in a live crime scene, not an isolated laboratory.

Malicious Actors (negative) This stakeholder group is implicitly cited in the justification for the use of deception in social network attack studies: if these stakeholders are aware of the defensive mechanisms, they can take action to avoid the detective mechanisms, including attacking the researchers or the accounts they are using for their experiments [10].

3.1.2 Primary Stakeholders

Users of the Social Network platform (positive) The benefit to users from developing a detective and protective mechanism to defend against malicious spamming is an increase in security. The primary risk to users of the platform is loss of privacy, however adding deception into the mix raises the potential for harm as noted in Section 3.3.

Law Enforcement (positive) As this research is focused on criminal activity, there is the distinct possibility that agents of law enforcement, who have the authority and responsibility for investigating and prosecuting criminal activity, may have an active investigation ongoing at the same time that researchers are interacting with the botnet. If researcher actions manipulate the crime scene and inject false information that misleads or hampers a criminal investigation, this can harm the efforts of law enforcement to do their duties and in turn, hurt the users who the researcher claims to be helping defend. The situation is entirely different if researchers perform experiments in an isolated offline environment where there can be no conflict, but to perform research using active and clandestine methods that alter a crime scene incurs real risk of harm. While human subjects review does not consider impact on law enforcement, researchers should still consider how their actions impact this stakeholder group in terms of reputational harm to themselves and their institution, should a serious conflict occur.
3.1.3 Secondary Stakeholders

Social Network Platform Owners (positive) The case studies just reviewed involve the Facebook and Twitter platforms. It is clear that there may be a benefit to these service providers if an effective detective and defensive mechanism can be developed to address attacks on their user base.

What is less obvious is the potential for harm to these service providers. The most direct potential harm relates to the Acceptable Use Policy (AUP), which we have seen in Section 1 gives users expectations about how the system is to be used by other users (which includes the researchers’ fake accounts). If these users feel that their rights have been violated, they may file complaints with the service provider, who must now expend resources to address. If a researcher is performing an experiment without the knowledge and involvement of the service provider, and the researchers’ actions (even final publication of results) make users aware of those actions and they file complaints about them, the service provider may suffer what they consider to be harm and could reasonably take action. This could include filing criminal or civil charges against the researcher. Deception, if not properly addressed, can add fuel to the fire of anger from users who believe their autonomy has not been respected.

Lastly, the problem of conflicts with law enforcement or with other researchers could both be mitigated through researchers engaging with social network platform owners and working with them, rather than simply performing potentially harmful research independently and in isolation.

Spammers and Criminals (negative) Negatively inclined stakeholders who profit from delivery of spam, theft of login or financial account credentials, etc., benefit when users’ defenses are low, or when researchers publish information about vulnerabilities without also publishing mitigation information, or following through on delivering defensive countermeasures once the general public knows of the researchers’ findings.

3.2 When does research become “Human Subjects” research

Researchers around the world have various restrictions and requirements for ethical review of their research when it poses harm to humans. In the United States, these restrictions and requirements are spelled out in the Code of Federal Regulations, 45 CFR 46 [7], also known as the Common Rule as it applies the same way to research funded by any federal agency. Ethical review in the U.S. is performed by bodies known as Institutional Review Boards (IRBs).

The most obvious activity that poses harm to humans is direct interaction during experimental medical procedures, but there is another risk of harm to humans involved in research that comes from collecting and using “identifiable private information.” The Common Rule includes in its definition of identifiable private information, “information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public.” This may include electronic communications such as postings in a social network that are restricted by privacy settings, geographic/geo-spatial coordinates associated with the use of a mobile device, or keyboard activity captured by malicious actors recording online financial transactions.

There is a distinction that can be made between studying humans who are using a social network (which could fairly clearly be seen as human subjects research), and studying malicious software attacking the users of the social network. One might argue that they are studying malicious software, not humans, therefore their research is not subject to IRB review. This may be a distinction without a difference, however, when it comes to the potential of harm to humans. This point is made not only by the Menlo working group members in the Menlo Report [5], but also by the Association of Internet Researchers (AoIR) in their ethics guide [15] and in guidance provided by SACHRP [1].

Collecting those communications thus falls under the “identifiable private information” category and likely triggers the requirement for IRB review. Even when an IRB approves a study, however, there can be problems. IRBs have limitations in their technical capabilities and may not fully understand the methods used to anonymize data, nor their limits [18]. One example is a 2008 study of taste preferences of students via their Facebook profile data, which was examined by Zimmer [24]. Not only did anonymization fail, allowing linkage of data back to students at the specific University, but the data was collected by research assistants who were members of group that was restricted to only students at that University. Those students believed their profile data, being part of a closed group, was not publicly available. Release of data identifying students in the United States, without their permission, is regulated by federal law and is spelled out as such in many university’s policy statements [12].
olation of this expectation of privacy was a key source of controversy with this study.

### 3.3 Ethical Issues Surrounding Deception

On top of the privacy issue is the issue of the use of deception. Deception in research studies involves “deliberately misleading communication about purpose of research and/or procedures employed’ [16].” By definition, when a person whose data is collected is deceived in some way, they are deprived of being fully informed about the nature of a research study, its benefits and risks, and are not allowed to make an autonomous decision about whether they want to be involved in the research or not. Deception raises issues of: Violating respect for persons by manipulating people to do something that they otherwise might not want to do; Violating the right to choose what to do based on relevant information; Possibly causing distress when it is later discovered that one was deceived. In short, “If use is not disclosed in advance, consent to research is not valid [16].”

The topic of deception is covered explicitly in the Menlo Report: “Sometimes informing stakeholders about the research procedure, purpose, risk-benefit analysis, and withdrawal opportunities impacts the scientific integrity of research results. Informing research subjects that some web sites are fake during a research experiment on phishing vulnerabilities could negatively impact the research validity by altering the subject’s behavior. Appropriate Respect for Persons in such deception research can typically be achieved by debriefing the subjects after the research is completed. Debriefing is typically required when deception is used in order to mitigate harm resulting from loss of trust in researchers by those subjects who were deceived [5].”

### 3.4 Waiver or Alteration of the Informed Consent Requirement

Obtaining informed consent is not always strictly necessary. 45 CFR 46.116(d) states, “An IRB may approve a consent procedure which does not include, or which alters some of the elements of informed consent [provided] the IRB finds and documents that: (1) the research involves no more than minimal risk to the subjects; (2) the waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) the research could not practicably be carried out without the waiver or alteration; and (4) whenever appropriate, the subjects will be provided with additional pertinent information after participation [8].”

If any personal data was collected, “debriefing should include offer to withdraw data [16].” This could partially be mitigated by not actually keeping any data that is processed in order to collect the malicious artifacts that are the true reason for monitoring social network posts using social honeypots. This should thus be clearly stated in the debriefing statement in order to minimize harm and respect persons.

### 3.5 Debriefing Statement or Request for Waivers

A researcher who wishes to request alteration or elimination of the requirement for obtaining informed consent must be able to explain to an IRB committee why a waiver is necessary and how they are going to protect humans involved in a research study who have not provided their informed consent. They may also be asked to debrief those persons who were deceived after the study is over. It seems logical to prepare for both of these situations by preparing in advance the kind of language that would go into a consent form or debriefing statement and modify it slightly for each specific use.

For example, a debriefing statement may contain language such as the following:

> We wish to inform you that your account was involved in a research study. The purpose of the study was to understand and counter the threats to users of social networks by people with malicious intent who send false messages intended to infect your computer, or steal your login credentials, bank account, or other very private personal identifiers.

> You should be aware that the investigators in this study have used fake identities and invited you to connect them into your social network. This was necessary in order to receive actual malicious communications sent through the social network. In doing so, the investigators have intentionally withheld any mention that these identities do not represent real people. There was no intention of collecting any personal communications or private information about you or others in your social network. You are not identifiable as an individual in any research results. Efforts have been taken to avoid or destroy any such information as soon as it was determined to not be hostile.

> The use of deception was necessary to also prevent malicious actors from being able to readily spot these identities and avoid them, thus preventing the researchers from observing how these malicious actors are preying
on you and other innocent users of the social network. If you have any questions or wish to confirm removal of any/all identifiable data, please contact the investigators at email.address@example.com.

Of course there is risk (i.e., a burden) to the researchers themselves when having to communicate even a small percentage of a million users of a social network around the world. This burden must be acknowledged by the researcher, not pushed aside as an inconvenience. As a community, the challenge of how to address issues of informed consent, debriefing, etc. at this scale is one we must also embrace.

4 Conclusions

We have very briefly looked at the subject of social honeypots and the risks involved in performing research on communications data in large social networks without consent of the users of those networks. We have also include the subject of deception in such research, and its implications.

This look is done from the perspective of academic research in the United States, which is (whether researchers know it, like it, or neither) is regulated by federal code and subject to Institutional Review Board review. Other researchers outside of this sub-population may have other ethical review regimes in which they operate.

We have seen how researchers of ethics in the biomedical and computer science fields view such potentially controversial research and what they recommend in terms of research design and communication with IRBs and research subjects.

We all want to help those who are being attacked by malicious actors through the internet via social media. At the same time, we need to respect those who we are trying to protect. We can do this by learning the ethical issues involved with our proposed research methods and help communicate to those we are trying to protect that we respect their autonomy and are doing everything in our power to minimize the harm they experience. In doing so, senior researchers will also foster a culture of ethical and responsible conduct of research for those junior researchers who follow.

References


Mock IRB Application for Social Network Honeypot Study

This mock IRB application illustrates how an actual IRB application might be crafted. The language and details were derived from the content of Zhu et al [SODEXO: A System Framework for Deployment and Exploitation of Deceptive Honeybots in Social Networks, 2012. http://arxiv.org/pdf/1207.5844.pdf], but was written without the involvement of that paper’s authors. The sole purpose is to provide a workable example based on details from an actual case study.

Overview of the study

This study implements an integrated system that can be interfaced with social networking sites for creating deceptive honeybots and leveraging them for gaining information from malicious botnets. A honeybot is a program that monitors communications to a false identity in a social network, rather than that of an actual human being. A malicious botnet is a set of these false identities that is constructed by a malicious actor and used to attack innocent users of a social network. The researchers are interested in how malicious botnet operators engineer these attacks and will use the honeybots to track and counter them.

The first phase of the study involves sending out deceptive friend requests to real individuals in order to constitute a realistic social network of over 1 million users and to monitor posts within that social network. The users would not be aware that the friend request was not genuine and they would interact with these fake users of the social network just as they would with other legitimate users. The second phase of the project will capture strategic interactions between honeybots (which monitor communications between the social network users and the false identity that invited them to join the network) and malicious botnets (which are attempting to infiltrate the users’ computers). The study will use quantitative methods to determine if honeybots can be used to design a protection and alert system to protect a social network from specially crafted false identities that look like real individuals. Once deployed, these honeybots monitor the message traffic of their neighbors and follow all links to find which ones attempt to infect users’ computers. When a malicious link is identified, the system follows command and control traffic to learn how the botnet is being controlled. There is no intention of collecting any personal communications or private information about users of the social network, nor of identifying any individuals in any publications or system output. Efforts will be made to avoid or destroy any such information as soon as it is determined to not be hostile.

Purpose of the study

The purpose of this study is to understand and counter the threats to users of social networks by people with malicious intent who send false messages intended to infect innocent users’ computers, or steal their login credentials, bank account, or other very private personal identifiers. This is accomplished by infiltrating malicious botnets using social honeypots designed to collect as much information as possible from the botnet. The intended outcome is to achieve a significant decrease in the infected population of a large social network.

Stakeholders

Key stakeholders involved in this study include researchers and malicious actors. Primary stakeholders include users of the social network platforms who are victims of cyber attacks and the agents of law enforcement investigating those crimes. Secondary stakeholders include the owners/operators of the social network platforms and criminal organizations using compromised hosts for criminal acts.

Recruitment

The subject of study is not the users of the social network, but instead the malicious software used to attack those users. The study thus does not recruit human subjects in the common sense, but rather interacts with users of an investigator-crafted social networks in order to monitor the communication traffic within the social network. The investigators in this study propose to use false identities and randomly invite users to connect to these false identities in order to create a social network of over 1 million accounts. Additionally, the investigators propose to use deceptive means to initiate these friend requests, so as to not allow the malicious actors to know that these “honeypot” accounts are monitoring and/or counteracting their criminal activities.

Procedures

A system that deploys honeypot identities first initiates a number of friend requests to randomly selected users from specially crafted false identities that look like real individuals. Once deployed, these honeybots monitor the message traffic of their neighbors and follow all links to find which ones attempt to infect users’ computers. When a malicious link is identified, the system follows command and control traffic to learn how the botnet is being controlled. There is no intention of collecting any personal communications or private information about users of the social network, nor of identifying any individuals in any publications or system output. Efforts will be made to avoid or destroy any such information as soon as it is determined to not be hostile.

Benefits

The primary benefits of this study are achieved by the second part of the proposed system, which takes advantage of the processed command and control traffic to generate signatures for malware, URL links, domains, or other indicators of malicious activity that can generate
intrusion detection signatures, blacklist spam filters, and user alerts. Owners/operators of the social network platform benefit from new defensive mechanisms to protect their users. Law enforcement agencies may benefit from insights into how social network botnets are commanded and controlled.

**Risks/Problems**

The principle risks are to users of the platform in the form of loss of privacy and the potential collection of identifiable private information. There is an additional risk due to the use of deception. A potentially large burden is placed on both the owners/operators of the social network platform and the researchers themselves when the approximately 1 Million users are debriefed after the study has been completed and informed that they were deceived by the friend requests to which they responded for non-existent individuals. There is also a potential burden placed on the platform providers who may expend resources investigating the false identities, believing they are being performed by negatively inclined malicious actors instead of positively inclined researchers.

**Consent Issues**

There are two primary stakeholder groups involved in data collection: the positively inclined user base of the social network, and the negatively inclined malicious actors who are attacking them. Any mention of the study in order to obtain informed consent risks rendering the research useless. It is therefore necessary for the investigators to request waiver of informed consent and permission to use deception to recruit users into the social network used for monitoring malicious infection and command and control traffic. After the research has been concluded, it would be possible to inform users that they were involved in a research study and explain how risks to their personal data and identities were mitigated.

**Debriefing Statement**

The following is proposed debriefing language:

We wish to inform you that your account was involved in a research study. The purpose of the study was to understand and counter the threats to users of social networks by people with malicious intent who send false messages intended to infect your computer, or steal your login credentials, bank account, or other very private personal identifiers.

You should be aware that the investigators in this study have used fake identities and invited you to connect them into your social net-