ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/228508220

Towards community standards for ethical behavior in computer security research

Article · July 2010

CITATIONS
20

READS
653

3 authors:

David Dittrich
Liminal Information Corp
37 PUBLICATIONS **1,683** CITATIONS
SEE PROFILE

Michael Bailey
University of Illinois, Urbana-Champaign
75 PUBLICATIONS **5,396** CITATIONS
SEE PROFILE

Sven Dietrich
City University of New York - Hunter College
44 PUBLICATIONS **1,053** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Machine learning and security View project

Ethics in Computer Security Research View project

All content following this page was uploaded by David Dittrich on 17 June 2014.

The user has requested enhancement of the downloaded file.

Towards Community Standards for Ethical Behavior in Computer Security Research

David Dittrich
University of Washington
dittrich@u.washington.edu

Michael Bailey
University of Michigan
mibailey@eecs.umich.edu

Sven Dietrich
Stevens Institute of Technology
spock@cs.stevens.edu

Stevens CS Technical Report 2009-1, (revised) September 25, 2009

ABSTRACT

Since the first distributed attack networks were seen in 1999, computer misuse enabled by *botnets*, *worms*, and other vectors has steadily grown. This rapid growth has given rise to a variety of ethical challenges for researchers seeking to combat these threats. For example, if someone has the ability to take control of a botnet, can they just clean up all the infected hosts? Can we deceive users, if our goal is to better understand how they are deceived by attackers? Can we demonstrate the need for better methods, by breaking something that people rely on today? When one considers the implications of something like botnet cleanup – the blind modification and possible rebooting of thousands of computers without their owners' knowledge or consent – this complexity becomes all the more obvious. To be effective, we must find ways to balance societal needs and the ethical issues surrounding our efforts, lest we drift to the extremes—becoming the very thing we deplore, or ceding the Internet to the miscreants because we fear to act. In this paper, we endeavor to build expertise in practical decision making, as well as to suggest a path towards development of community standards and enforcement mechanisms governing basic and applied computer security research.

1. INTRODUCTION

Modern threats such as Denial of Service Attacks, Worms, Viruses, Phishing, and Botnets underscore the need for security research in an increasingly networked and computationally reliant society. Unfortunately, as our understanding of these phenomenon have grown, so has the uncertainty in the computer security research community on the appropriate ways in which to observe and address these problems.

For example, consider the area of botnet research, which centers around the detection and mitigation of large numbers of infected hosts, or *bots*, networked into a single distributed system, or *botnet* [65]. We have recently seen a steady increase in the amount of criminal activity using botnets, and along with this has come an increase in the number of academic research and federal funding to counter the botnet threat. This criminal activity is compounded by the emergence of politically motivated attacks, such as those against elements of the cyber-infrastructure of Estonia. Responses to these threats are varied, from passive measurement and observation, to calls for the legal right to defend computer systems from attack using aggressive countermeasures [58, 77].

Unfortunately, the structured public discussion of an ethical framework to guide decision making about actions taken while researching and countering botnet attacks, and indeed in a broader set of computer security research, has not kept pace. As a result, we left with uncertainty and inconsistency both in how we make personal decisions and the feedback we get from peers.

Existing structures for determining ethical behavior (e.g., Institutional Review Boards (IRB), military and intelligence rules, and Professional Codes of Conduct) fail to provide detailed actionable guidance due to many reasons: the absence of technical expertise in this specific domain; the lack of authority over the research; and/or a lack shared community values [6]. There is growing frustration expressed by researchers, program committees, and professional organizations about the limits of ethical research and who has responsibility to enforce them [6, 31].

Our primary goal in this work is to build expertise in practical decision making by illuminating ethical issues and analytic tools, and showing how they may be applied using case studies. Secondarily, we suggest how policies might be formulated through community consensus and how policy enforcement bodies (e.g., program committees, IRBs, grant funding agencies, or ethics boards) may use these policies in their deliberations. To help achieve these goals, this paper provides:

- **An Exploration of Existing Ethical Arguments**. We are certainly not the first authors to grapple with the notion of ethics in general nor ethics in an computer society. Existing work in this field can help us narrow the scope of our efforts and provide guidance on building *consistent* and *coherent* arguments for ethical principles.
- **An Example Framework for Security Research**. We create an amalgam of existing approaches to human subjects research, professional principles, and active response justification in order to create a quantitative framework for judging risk and benefits in computer security research.
- **Exploration of Ethics through Case Studies**. While our framework explicitly does not draw conclusions about when a piece of research is ethical or unethical, it highlight the relevant ethical issues the research raises. We review 25 recent case studies and apply the framework to a significant fraction of these studies.

2. ETHICS, LAW, AND COMMUNITY STANDARDS

The study of ethics has a long history. While comput-

ing and the Internet provide recent twists to long debated ethical issues, the study of even these new applications is a field unto itself. In this section, we provide a context for the remainder of our work by examining the fields of ethics, law, various definitions of community, and existing standards of behavior.

2.1 What is ethics?

Ethics is often defined as a set of morals or guiding principles intended to govern the conduct of individuals and groups within a population (i.e., a profession, a religion, or society at large.) The definition of *computer ethics* has various interruptions in line with this broader definition, and several are explored in Bynum and Rogerson [14]. One of the most often cited of these is from Moor [57]:

> A typical problem in computer ethics arises because there is a policy vacuum about how computer technology should be used. Computers provide us with new capabilities and these in turn give us new choices for action. Often, either no policies for conduct in these situations exist or existing policies seem inadequate. A central task of computer ethics is to determine what we should do in such cases, i.e., to formulate policies to guide our actions. Of course, some ethical situations confront us as individuals and some as a society.

Note that we agree with others in the field who argue that these polices, once developed, are neither absolute laws, nor complete frameworks, nor checklists to be followed blindly [14]. They are never likely to be complete nor the polices mutually exclusive. As such our approach here is close to that of Johnson and Miller [46] in that we are concerned with building expertise in practical decision making. Theoretic ethics and ethical systems are useful in these processes, but not ends in themselves.

2.2 Law versus ethics

The law is in some ways a set of norms that are written to guide behavior within a society. These legal norms can codify another set of moral and ethical norms that are generally agreed upon by that society. These sets of norms are not, however, the same. For example, we may agree that lying to a friend is unethical, but lying to a friend is not always illegal. Lying under oath, on the other hand, is always illegal. In relation to security research there may be many laws in many countries that are implicated by a given action taken by a researcher. But what does this have to do with ethics? Both legal and ethical considerations matter to security research in several ways.

- Adherence to ethical principles may be required to meet regulatory requirements surrounding academic research (e.g., IRB review and NSF grant requirements.)
- An ethics-based decision making framework may inform academics, security professionals and amateur security researchers as to how to decide on actions to take in response to a criminal botnet.
- It can illuminate the line between beneficial acts and harmful ones.
- It can describe all parties involved, their rights and responsibilities, and how to resolve conflicts between competing interests.
- What may be most important in terms of reputation is being able to clearly justify one's actions should those actions come into conflict with the law, or generate public controversy.

Developing a workable ethical framework is only a first step, however. Having guidelines that embody a set of norms accepted within the security field improves the decision making process. It gives the public a sense of security in knowing that individuals are acting in the best interest of society. Once these norms are accepted, they can then be considered and adopted within the legislative process to advance the common law.

This is similar to the field of *computer forensics*, where the issue of the admissibility of scientific evidence in trial is concerned. Based on standards established in a Supreme Court case in 1993, known as *Daubert*, [64] courts will accept testimony involving computer forensic evidence if it meets criteria of (a) relevancy, and (b) reliability. It is the second criteria that matters for this discussion. The court suggested that judges evaluate testimony for *scientific validity* and ensure its proper application to the facts of the case, saying:

> Many considerations will bear on the inquiry, including whether the theory or technique in question can be (and has been) tested, whether it has been subjected to peer review and publication, its known or potential error rate and the existence and maintenance of standards controlling its operation, and whether it has attracted widespread acceptance within a relevant scientific community. The inquiry is a flexible one, and its focus must be solely on principles and methodology, not on the conclusions that they generate.

If the computer security research community develops ethical principals and standards that are acceptable to the profession and integrates those as standard practice, it makes it easier for legislatures and courts to effectively perform their functions. If the broader society also accepts those principals and standards, an even greater benefit results in terms of societal trust in computer security research.

While both influenced by ethical frameworks, and serving as a guide for classes of ethical behavior, a full discussion of the legal issues surrounding computer security research is beyond the scope of this paper. Interested readers are encouraged to examine an increasing body of work in this area. [13, 66, 50, 28]

2.3 What do we mean by community?

In this paper, we use terms like *researcher*, *organization*, *community*, and *society*. These terms apply in two primary contexts: the population taking action, which will use the ethical framework we suggest to guide decision making, and the population implicated in those actions and/or the intended beneficiaries of those actions (e.g., owners of bot infected computers.) The actor populations are, of course, also, directly or indirectly, members of the protected and beneficiary populations. We find three such populations of interest:

- **Individuals in Society** These are members of the general public who are independent security researchers

or computer hobbyists who are interested in computer security. This population has the least control of their actions, especially when harmful acts are not covered by criminal or civil laws. Laws do, however, still cover certain actions they may take. In terms of benefits, individuals in society (which encompasses the other populations below) are the principle beneficiaries of computer security research.

- **Professional Community** These are professionals who have roles that involve them in computer security research or incident response as part of their normal job duties. For example, network operators, security operators, forensic analysts, reverse engineers, computer security incident response staff, etc. Control over this population is principally governed by their employers' administrative policies, agreements signed with employers or clients (e.g., non-disclosure agreements) and contract terms. Harmful acts are punished by dismissal, disbarment, or legal actions.
- **Academic Community** These are people who have academic roles in educational institutions, primarily research staff, research faculty, and students (both undergraduate and graduate level) who are studying information security related topics. Control over this population can include both legal restrictions and institutional policies. Harmful acts would be punished by academic sanctions, dismissal, and/or legal actions.

Note that our notion of actors and beneficiaries here are meant to focus this work beyond many of the more general discussions that dominate existing computer ethics work. For example, many texts emphasize the role of general users of technology play in changing how we think about fundamental issues in society such as privacy and intellectual property rights [7]. Some work does focus on the discussion to the role of professionals involved in the application of this technology [14], but often the notion of professional is limited to the roles of software engineers and engineering mangers rather than security researchers. It is our intention in this paper to focus on the most specific of these actor populations: the academic security community.

2.4 What standards exist for guiding ethical behavior in our community?

The security community already has some standards and regulatory requirements to adhere to certain research protocols. In this next section, we examine some of these standards.

2.4.1 Rules of Engagement

In terms of analogues, cybersecurity is often compared with physical conflict (i.e., war.) As a result, the discussion may focus on the ethics of *responding* to computer attack in terms of *use of force* alternatives under theories of the *Law of War*, or *Law of Armed Conflict* (LOAC). For the purposes of this discussion, there are parallels to concepts embodied in the LOAC. For example, the LOAC requires *military necessity* as a pre-requisite for the use of force. It requires *distinction*, that is, actions must be directed against lawful combatants and military targets, not against civilians and civilian infrastructure. Lastly, the LOAC requires *proportionality*, that is a use of force less than or equal to the original harm or violation. As a result of international agreements and protocols, such as those defined in the Geneva Conventions, [39] militaries around the world operate under strict *Rules of Engagement* (ROE). These ROE guide decision making on the field of battle to ensure the actions of military personnel do not result in potential war crimes charges.

Yurcik [78, 79] discusses ethics in relation to attack and retaliation using *information warfare* (IW) tactics, and considers whether the lethality of IW operations affects the ethics of employing such operations in defense. This applies to military responses to attacks at the nation-state level, but sets the stage for the equivalent considerations of responses in non-military settings. Yurcik next considers *hack-back*, or aggressive responses to computer attack by attacking back. [44] More complete analyses of the application of international law and the law of war to state-directed IW – also known as *cyberwarfare* – operations were done by Sharp [70] and Wingfield. [76] The actors here are primarily nation-states, not individuals.

Dittrich and Himma [25] discuss the legal and ethical frameworks for responding to computer intrusions. Their research identifies three ethical principles as being central to consideration of aggressive counter-measures: the *Defense Principle*, the *Necessity Principle*, and the *Evidentiary Principle*. Dittrich and Himma build on previous work by Yurcik, specifically focusing on the non-military considerations for response, as well as considering transition of response from civilian to military realms. Himma later expands [36] on previous work with Dittrich to include the *Punishment Principle* and the *Retaliation Principle*.

2.4.2 IEEE, ACM, and other professional standards

The Association of Computing Machinery's (ACM) *Code of Ethics and Professional Conduct* [5] consists of three distinct parts which highlight fundamental ethical considerations, specific professional responsibilities, and leadership imperatives. Section 1 entreats members to: "contribute to society and human well-being" (Section 1.1) and to "avoid harm to others" (Section 1.2), along with six other principles (e.g., don't discriminate, be honest, respect privacy). Professional responsibilities include calls that "ACM members must obey existing local, state, province, national, and international laws unless there is a compelling ethical basis not to do so," (Section 2.3) and to "access computing and communication resources only when authorized to do so." (Section 2.8), along with maintaining competence, accepting review, etc.

The IEEE also maintains the *IEEE Code of Ethics* [40], which, although more abbreviated than the ACM version, contains many of the same imperatives. Specifically, the code commits members "to the highest ethical and professional conduct." Members agree to avoid conflicts of interest, be honest, engage in responsible decision making, accept criticism of work, etc. Of particular interest are the mandates, "to improve the understanding of technology, its appropriate application, and potential consequences," and, "to avoid injuring others, their property, reputation, or employment by false or malicious action."

These are certainly not the only such codes of conduct for computer professionals. For example, IEEE and ACM have approved a joint Software Engineering Code of Ethics [33] and there are numerous professional organizations with codes whose headquarters are outside the United States (e.g., the

Institute for the Management of Information Systems in the UK, [41], Australian Computer Society, Canadian Information Processing Society (CIPS)). In addition some individual companies and academic institutions have their own ethical codes (e.g., Gateway, Texas Instruments, University of Virginia, Howard University), but these are by no means universal.

There are also professionals forming affinity groups, often described as *mitigation communities*, for the purpose of analyzing and responding to distributed malicious attacks, often called *botnets* for short. One such group, a non-profit organization called the *Shadowserver Foundation*, describes the *Standards and Guidelines* for how the group operates on their web site [3]. In the process of their malware analysis, they tend towards using *passive* methods of analysis that do not extend beyond what the malware itself does as part of its normal operation (i.e., simply participating in a command and control channel) and they try to ensure they do not harm infected computers. For example, they do not interact with the malicious actors controlling the botnet, nor do they attempt to execute commands to "clean up" bot infected hosts as was done by BBC reporters in *Case* 10 . In terms of the Levels described in the Active Response Continuum, [25] Shadowserver Foundation prefers activities at the lower level of aggressiveness (*Benign* to low *Intermediate*) within Level 4 (*Uncooperative* response), while also working with affected sites whenever possible at Level 3 (*Cooperative* response) to help those sites clean up their own (or their customers') systems.

2.4.3 Human Subjects Standards

Biomedical and behavioral research in academia (in the United States) is bound by regulatory requirements to examine certain ethical considerations related to the protection of human subjects. In response to a number of incidents of medical research being performed on individuals without their knowledge or consent, the National Research Act was passed in 1974. These incidents included syphilis studies involving low-income African-American males in Tuskegee, Alabama in the 1930s, and medical experiments performed on prisoners of war in World War II (protection of whom was mandated in the Nuremberg Code following Nazi war crimes trials.) This act established a regulatory body, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. In 1979, the National Commission prepared a document known as *the Belmont Report*. In 1981, the Department of Health and Human Services (DHHS) released a regulation (45 CFR Part 46, Subpart A) based on the Belmont Report, known simply as the *Common Rule*. Common Rule defines requirements for *research* involving *human subjects* that apply to individual researchers, their institutions, and their related Institutional Review Boards (IRBs).

The National Research Act was also informed by the Declaration of Helsinki (1964, revised several times since). This declaration addresses issues of research protocols involving humans in terms of risks and benefits, informed consent, qualifications of researchers, etc. While the National Research Act applies to those in the United States, the Declaration of Helsinki similarly informed a set of standards applied to clinical research around the world known as *Good Clinical Practices* (GCP.)

The three basic ethical principles and their application described in the Belmont Report are:

- **Respect for Persons** Participation as a research subject is voluntary, and follows from informed consent; Individuals should be treated as autonomous agents, whose right to decide about their own best interests is to be respected; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
- **Beneficence** Do not harm; Maximize possible benefits and minimize possible harms; Systematically assess both risk and benefit.
- **Justice** To each person an equal share in treatments and benefit of research according to individual need, effort, societal contribution, and merit; There should be fairness of procedures and outcomes in selection of subjects.

One of the key elements of *respect for persons* is the concept of *informed consent*. Researchers are required to clearly and in plain language describe the research activities in which the subject will be involved, and the risks and benefits of participation in research. The researchers must also obtain affirmation (usually in writing) that the subjects are aware of the risks/benefits, that their participation is entirely voluntary and separate from any other activities (e.g., normal medical treatment they may be receiving at the time), and they agree to participate.

2.4.4 Internet Activities Board Standards

Engineering and best practice standards for the internet are defined by documents approved by the Internet Engineering Task Force (IETF.) They are known collectively as Request for Comment (RFC) or Best Current Practice (BCP) documents, and each is numbered uniquely. RFCs are also used as informational documents that do not necessarily specify standards, but are officially sanctioned and maintained by the IETF. Two such RFCs authored by the Internet Advisory Board (IAB) involve ethics in relation to measurement activities, research, and general internet use.

RFC 1087, "Ethics and the Internet," is a general policy memo that "endorses the view of the Division Advisory Panel of the National Science Foundation Division of Network, Communications Research and Infrastructure" in characterizing unethical behavior that involves unauthorized access, disruptive or wasteful activity, or compromise of user privacy [9]. The bottom line is that internet users – which includes researchers performing experiments – are responsible for their own actions and should behave in a constructive, rather than destructive, manner for the good of all users of the internet.

RFC 1262, "Guidelines for Internet Measurement Activities," is an informational document that stresses it is important "that data collection activities do not interfere with the operational viability and stability of the network, and do not violate considerations regarding privacy, security, and acceptable use policies of the network. [16]" The IAB suggests that researchers attempt to "alert relevant service providers using mechanisms such as bulletin boards, mailing lists and individual mail communications." They also suggesting making information about research methods publicly available "by anonymous FTP or other means" and/or by informing Carnegie Mellon University's Computer Emergency

Response Center (CERT, now known as the CERT Coordination Center, or CERT/CC) in advance of experiments, in order to allow remote sites to differentiate benign research from break-in attempts. A list of specific conditions that researchers are suggested to carefully consider and meet in developing experimental methodologies is provided.

While the guidelines in RFC 1262 may have been appropriate and easily followed by researchers and involved sites in 1991, and the network described by RFC 1087 was a "national facility [under the fiduciary responsibility of its] U.S. Government sponsors" in 1989, the internet has long since outgrown its original research-centric roots and the volume of malicious activity has grown with it. Much of the guidance (e.g., notification of experiments via bulletin boards or anonymous FTP sites, or manual detection and/or vetting activity by asking CERT/CC if they were informed of an experiment taking place) is no longer practical. However, the general advice concerning evaluation of issues of integrity, availability and confidentiality of data, and careful consideration of risk/benefit comparisons, is just as appropriate today.

2.4.5 *Proposed Intelligence Community Standard*

Two former members of the U.S. intelligence community, in an article in the Communications of the ACM, [67] propose an ethics code for U.S. intelligence officers. Their primary goal is to help reduce the number of bad decisions that may be made in the heat of the moment, and to offer some protection against public outrage when and if classified actions become publicly known. They believe a code of ethics could, "help citizens understand the rationale for [given actions] and lesson adverse reactions or possibly offer an opportunity to further refine the language (and constraints on future actions) to be more in line with national values that may change over time." They urge a code that (a) defines behaviors to aspire to, (b) is defined in simple and easy to understand language, (c) reflects such issues as "lawfullness, transparency, accountability, truthfulness, examining consequences of plannaed actions, and protection of innocent individuals." Their primary motivation in offering the code is to, "foster discussion, deliberation, and debate that would help people internalize the code" through reflection, introspection, and contemplation during training.

The code itself is described in a side-bar as follows:

> 1. First, do no harm to U.S. citizens or their rights under the Constitution.
> 2. We uphold the Constitution and the Rule of Law; we are constrained by both the spirit and the letter of the laws of the United States.
> 3. We will comply with all international human rights agreements that our nation has ratified.
> 4. We will insist on clarification of ambiguities that arise between directives or law and the principles of this code. We will protect those within our institutions who call reasonable attention to wrongdoing.
> 5. Expediency is not an excuse for misconduct.
> 6. We are accountable for our decisions and actions. We support timely, rigorous processes that fix accountability to the responsible person.
> 7. Statements we make to our clients, colleagues, overseers and the U.S. public will be true, and structured not to unnecessarily mislead or conceal.
> 8. We will resolve difficult ethical choices in favor of constitutional requirements, the truth, and our fellow citizens.
> 9. We will address the potential consequences of our actions in advance, especially the consequences of failure, discovery, and unintended or collateral consequences of success.
> 10. We will not impose unnecessary risk on innocents.
> 11. Although we may work in secrecy, we will work so that when our efforts become known, our fellow citizens will be proud of us and of our efforts.

While there is no strict equivalency between the constitutionally authorized activities of intelligence community agents, there are some similarities in the motivations and desired outcomes described in this paper and in Snow and Brooks. [67] Most importantly is the desire for actions taken in private to be acceptable if made known to the general public (whose computer systems are implicated in many cases.)

2.5 Limits on existing standards?

Allman [6] discusses the potential role of conference Program Committees (PCs) in guiding researchers in terms of the ethical foundations for their research methodologies. The ACM code is cited as one guide that PCs may apply in judging academic papers submitted to them for review, however Allman mentions that interpretations can be varied and application of the code to specific actions difficult. One could read the ACM code, Section 2.3, and apply the ethical principles cited in this paper and conclude some research is acceptable, while another could read its Sections 1.1 and 2.8 and conclude the same research is unacceptable. Allman also mentions IRBs as potential arbiters of the ethics of botnet research, but points out that IRBs deal only with research involving human subjects, historically from the fields of biomedicine, psychology, etc. Allman also questions whether IRBs share computer security researchers' value systems, or have sufficient domain expertise, to judge the risks involved in computer security research. This is consistent with some of the concerns cited by Garfinkel [31].

For example, consider botnet research. In the medical research context, the research subjects themselves are the eventual beneficiaries of the research outcomes (and to a larger extent the rest of society in general.) In the security research context, the research subjects are often criminals and their tools, which happen to involve (most often unknown to their owners) the computers of innocent third parties. This means there are two potential sets of beneficiaries who potentially have an inverse benefit/harm relationship to one another. That is to say, publication of some research results may have a small benefit to society in general, while the criminals whose tools are the subject of research may have a much larger benefit. The criminals may learn how to improve their attacks, or make them harder to detect and mitigate. This is especially true of publication of theoretical research that postulates new and more potent types of malicious software, which could serve as a blue-print for criminals. This is a very complex calculus that sometimes

involves initial non-public disclosure of research results, and very carefully timed public disclosure, in order to assist law enforcement or provide lead time for security operations elements to act (e.g., in cases of vulnerability disclosure [60] Non-public disclosure is diametrically in opposition to typical academic research, where "publish or perish" and "open access" are common mantras.

As another example, consider the works of Denning and Spafford who discuss ethics in the context of those engaged in *computer intrusions*. Denning [21] describes the opinions of hackers who were interviewed about computer intrusions as to whether those acts were ethical or not. Some hackers believed that certain malicious actions where wrong and unethical (e.g., "breaking into hospital systems," "reading confidential information about individuals," "stealing classified information," "committing fraud for personal profit.") Some hackers believed that exploring computer systems was ethical, provided that "the objective is to learn and avoid causing damage." Spafford [69] looks at similar acts in terms of *right and wrong*, and whether a *greater good to society* is achievable by computer intrusions. In Spafford's analysis, computer intrusions may only be ethically justifiable in the most extreme cases, such as to save a human life in an emergency. In discussing publication of worm or virus code (which may be capable of resulting in harm to innocent third-parties) he states that, "publication should serve a useful purpose; endangering the security of other people's machines or attempting to force them into making changes they are unable to make or afford is not ethical."

Unfortunately, there is no commonly accepted framework within which decisions can (relatively) clearly and consistently be made. Nor is there currently an accepted venue in which to consider them. Should PCs be the arbiters, as Allman suggests, [6] or is that venue too closed? Both Allman and Garfinkel [31, 6] suggest Institutional Review Boards may have a role, but are they currently capable of making judgments about the issues raised in this paper, or do they even have oversight responsibility in enough cases that matter? (E.g., out of all the case studies cited in Section 3, an IRB might only judge Cases 4, 6, and 26 to fall into the category of *research involving human subjects* that would necessitate IRB committee review.) If the role of IRBs is to ensure compliance with the National Research Act, "to put a stop to researchers saying 'Trust me'," [31] does this support Burstein's suggestion that researchers, "participate in [legislative] reform efforts... to make known how the lack of a research exception affects them" and their research? [13]. Sicker, et al, [66] offer reasons such legislative reform is neither a timely nor especially effective solution and suggests that prosecutorial discretion may preclude the need for legislative reform (although this has its own risks.). Should computer security researchers be required to receive training similar to the *Education on the Protection of Human Subjects* mandated by NIH, and if so, what should be covered? How are these issues dealt with internationally?

While the limitations on scope, expertise, and lack of consensus are broader than a single discussion or single work, it is clear that the answers to these and other questions will require community dialogue and effort. For such a dialogue to be successful, we will need to draw from a rich set of experiences and build consistent and coherent arguments for the ethical or unethical behaviors contained therein. In the next section, we describe our efforts in building such a dialogue.

3. CASE STUDIES

We will now look at several case studies both inside and outside of the academic research setting, in terms of the ethical principles mentioned earlier. Not all of these are research specific, but all serve to illustrate the ethical questions involved.

3.1 Participating, Observing, and/or Breaking Something to Understand How It Works

Shining Light in Dark Places: Understanding the ToR Network. [55] [*Case* 1] McCoy et al. participated in the Tor network to analyze the types of traffic, countries using Tor, and possible abuses of the network. By running a modified Tor server, they were able to observe all traffic either being relayed (they were a relay for two weeks) or exiting the network (they were an exit node for another two weeks). Fully aware that the payload collection would be a problem, they tried to limit the amount of payload data being collected in the experiment. The main purpose of the work was one of discovery and measurement, and how to possibly limit the exposure of sensitive data, as they devised a method to detect logging by malicious routers. However, suggestions for improving and fixing Tor also emerged from this paper.

Learning More About the Underground Economy: A Case-Study of Keyloggers and Dropzones. [37] [*Case* 2] In order to study impersonation attacks, typically achieved using keyloggers, Holz et al. used a malware analyzer to locate so-called dropzones within malware samples. These dropzones are the places where keylog information gathered from the users is sent by the malware, to be later retrieved by the malware operators. At these dropzones, the researchers discovered 33GB of data from 173,000 compromised computers, containing 10,000 bank account and 149,000 e-mail account passwords. This study was conducted over a period of seven months in 2008 and aimed to study the underground economy and to automate the analysis process. The collected data was eventually handed to AusCERT, which acted as a notification broker for the victims.

Your Botnet is My Botnet: Analysis of a Botnet Takeover. [*Case* 3] Stone-Gross *et al.*, [72] at University of California, Santa Barbara, analyzed the Torpig botnet by taking control of the botnet for a brief period starting January 25, 2009. This was accomplished after reverse engineering the bot.

This reverse engineering allowed the research team to ultimately take over the botnet. They did this by identifying which domains it would generate on particular days, finding two domains (one in .com and one in .net) that had not yet been registered by the attackers, and registering them first. They then purchased web services at collocation providers known to be unresponsive to abuse complaints, and set up their own command and control (C&C) servers. The attackers eventually updated the bots, effectively taking control back ten days after the start of the experiment.

While they controlled the botnet, the researchers captured over 70 GB of data collected by the bots (8.7 GB of Apache log files, and 69 GB of pcap data.) This data was rigorously analyzed. It included credentials for 8,310 accounts at 410 different financial or commercial institu-

tions. 297,962 unique account login credentials were found. Passwords were analyzed for uniqueness (28% of the victims reused passwords) and Unix hashes were generated for these passwords, which was then fed through a popular password cracker (more than 40% of them were broken in less than 75 minutes.) 6,542 captured personal messages were identified that were (a) written in English and (b) longer than 250 characters. Keywords were used to categorize these messages, which identified topics of conversation including exam preparation, professional advice from lawyers and doctors, job seeking, discussing money or sports, exchanging insults, and looking for sex or partners online.

The researchers were able to uniquely identify the Torpig bots over time, getting around a fundamental problem in accurately counting bots. [63, 49] They observed over 180,000 infections out of 1.2 million total IP addresses observed, including several suspected instances of bots running in virtual machines that were assumed to be other researchers or individuals probing the botnet.

The actions taken by the researchers were justified using two ethical principles:

1. "The sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized, and
2. The sinkholed botnet should collect enough information to enable notification and remediation of affected parties."

The modified C&C server responded to bots with replies that kept them from moving off to attacker-controlled C&C servers, and no attempt was make to disable the bots by feeding them blank configuration files (avoiding potential unforeseen consequences.) Data collected from infected hosts was turned over to ISPs, law enforcement agencies, and the Department of Defense, leading to suspension of other domains actively being used by the attackers.

Why and How to Perform Fraud Experiments. [43] [*Case* 4] In this work the authors discuss their experiences with conducting fraud experiments (i.e., phishing). In particular, they focus on two studies: one in which they explore the impact on phishing source (i.e., someone trusted versus someone random) [42] and one in which they explore the impact of cousin domains (i.e., those which sound similar to the real domain) [30]. The purpose in this article was not to explore these studies in depth, but rather to highlight three important ethical issues associated with conducting these experiments. The first of these issues, that of informed consent, centers around whether it is ethical to not allow the participants to choose whether to participate in the study. Here the ethical considerations of the value from the study must be weighed against the fact that the study results change if the users know it is happening. A similar set of arguments are used in discussing the next issue, that of explicit fraud. As mentioned in the article, lying to users must be done with the utmost care, be overseen by a full IRB committee (i.e., not expedited review by a subcommittee), and should generally be avoided by researchers. Finally, the authors explore the notion of debriefing, that is, informing users after the study that they participated without their knowledge. Debriefing is generally a requirement when informed consent is waived as participation in research studies is voluntary. In this case, the IRB granted waiver of informed consent and debriefing based on risk/benefit assessments that were not articulated in the summary article.

Measurements and mitigation of peer-to-peer-based botnets: a case study on storm worm. [38] [*Case* 5] In April 2007, Holz, et al, performed Storm botnet enumeration experiments in which they infiltrated the Storm botnet and used features of the distributed hash table (DHT) that is used by Storm to enumerate the bots. They were able to observe the effect of other researchers who were simultaneously doing their own enumeration experiments, and specifically noted UCSD and Georgia Tech (among other unnamed sites) as being observable participants in the Storm botnet. They discuss two attacks – eclipsing, or *Sybil attack*, and poisoning – that could be performed to degrade or render inoperable the Storm botnet. Both could be argued to be positive outcomes. While not stated by Holz, these two attacks would also not have negative affects on the owners of compromised computers. While potentially disabling the botnet, at least temporarily, these attacks do nothing to help mitigate the botnet by assisting in cleanup efforts of individually compromised hosts.

Spamalytics: an empirical analysis of spam marketing conversion. [48] [*Case* 6] Kanich, et al, (2008) performed a study of the conversion rate of spam campaigns. Their analysis was achieved by infiltrating the Storm botnet and manipulating spam being relayed through systems they controlled by altering command and control (C&C) traffic, and using a fake web site that looked like web sites advertised by those responsible for setting up the illicit Storm botnet. The ethical considerations used to justify their experiments follow the principle of the use of *neutral actions* that *strictly reduce harm.* This was the first time research was performed to learn the conversion rate of spam campaigns. Alternative actions that *could also result from manipulation of C&C traffic*, which may result in an equal or greater moral good to society, are not discussed.

Studying Spamming Botnets Using Botlab. [45] [*Case* 7] John, et al, (2008) researched spam-generating botnets through analysis of email messages identified by email filters at the University of Washington (UW). Through the use of a botnet monitoring architecture incorporating malware analysis and network behavioral analysis, they were able to develop several functional defenses. be explicit about the risks that result from doing behavioral analysis of malicious botnets, and conclude that, "a motivated adversary can make it impossible to conduct effective botnet research in a safe manner." Observing that an attacker could design even benign looking C&C traffic that could result in the researchers' bots causing harm to third-party systems, they chose to be conservative and halted all network crawling and fingerprinting activity that would identify new malware binaries. They also stopped allowing any outbound connections to hosts other than a small set of known central C&C servers, which meant they halted all analysis of Storm (which uses variable ports for its obfuscated C&C servers.) By taking a very conservative stance, they are minimizing potential harm yet simultaneously limiting their future ability to do beneficial research.

P2P as botnet command and control: a deeper insight. [24] [*Case* 8] In 2006, Dittrich and Dietrich, began analyzing the Nugache botnet. Nugache, the first botnet to successfully use a heavily encrypted pure-P2P protocol for all command and control, was nearly impossible to observe through passive monitoring of traffic flows from the point-of-view of local networks. After fully reverse engineering the Nugache P2P protocol, a crawler was written that took advantage of weaknesses in the P2P algorithm. Several enumeration experiments were performed with the crawler, carefully crafted to ensure minimal impact on the botnet. This crawler, and the enumeration experiments performed with it, are similar to later efforts to enumerate the Storm botnet. [49, 38] The authors cite two key issues with botnet enumeration experiments: *accuracy* in counting, and *stealthiness*. They note the potential for researchers doing aggressive enumeration experiments to inflate counts obtained by other researchers, to hinder mitigation efforts, or to impede law enforcement investigations. Source code to the enumerator, and certain sensitive results from reverse engineering analysis, have not been made public to minimize the potential for malicious actors to leverage non-public knowledge for their own purposes. That could result in an imbalance of benefit to malicious actors over society at large, and potentially increase harm to society.

3.2 "Hack-back" and Aggressive Response

DDoS attacks against South Korea and U.S. government sites. [*Case* 9] Starting on July 4, 2009, web sites in the United States and South Korea were targeted by sustained DDoS attacks. Because these attacks were directed at government agency web sites, this matter drew immediate press attention and concerted efforts to mitigate the attacks.

On July 12, 2009, the organization Bach Khoa Internetwork Security, centered at the Hanoi University of Technology (HUT), announced publicly through their blog that they received a request for assistance from from the Korean CERT (KrCERT) and information that allowed them to identify 8 command and control (C&C) servers for the botnet suspected of performing the DDoS attacks [26]. BKIS claimed they "fought against C&C servers [and gained] control" of 2 systems located in the United Kingdom from which they remotely retrieved log files allowing BKIS to count and geolocate over 160,000 IP addresses around the world participating in the botnet. KrCERT believed the BKIS statement to falsely suggest direct involvement and complicity of both KrCERT and the Asia-Pacific CERT (APCERT) in potential violations of Vietnamese and international laws. KrCERT lodged an official complaint against BKIS with the Vietnamese CERT (VNCERT), who were themselves unaware of BKIS' activity prior to the BKIS blog posting. [61] VNCERT sent a letter to BKIS that was later made public via the VNCERT web site, "generating fierce debates on many online forums," and prompting BKIS to consider filing a law suit against VNCERT for defamation. [68]

BKIS cited a July 10, 2009, email from KrCERT to members of APCERT asking for urgent assistance in discovering the source of the DDoS attack as justification for taking the actions they did, and denied doing anything illegal. A BKIS representative stated that they used common tools and practices to discover the vulnerable C&C servers, and that accessing those systems remotely "doesn't require anyone's permission and anybody can do it." BKIS justified not reporting to VNCERT during the 2-day period of investigation by citing Article 43 of the Vietnamese government's Decree 64/2007, which states: "In urgent cases which can cause serious incidents or network terrorism, competent agencies have the right to prevent attacks and report to the coordinating agency later."

VNCERT claimed that, "BKIS should have only made its findings known to parties involved," and that "BKIS had made the announcement before confirming its findings." [68] This opinion conforms with the vulnerability disclosure policy published on the BKIS web site. This policy describes a much longer time frame and more deliberate steps for non-public reporting that could have prevented some of the controversy that arose. Not mentioned in cited articles was the potential that disclosure of the UK network block containing the C&C servers might hamper law enforcement investigation of a solid lead in what BKIS themselves called, "[an] urgent case, which could threaten the world." [61] The Vietnamese Ministry of Information and Communication finally became involved, believing both BKIS and VNCERT to have made mistakes, and requesting they both quickly resolve issues involving foreign parties and be more careful in future.

BBC TV: Experiments with commercial botnets. [51] [*Case* 10] In March 2009, the British Broadcasting Company (BBC) *Click* technology program chose to perform an experiment. Unlike the situation in *Case* 14 , direct control of the botnet was exercised. The BBC staff purchased the use of a malicious botnet identified after visiting internet chat rooms. They used that botnet for several purposes: (1) They sent thousands of spam messages to two free email accounts they set up on Gmail and Hotmail; (2) They obtained permission to perform a distributed denial of service attack against a site willing to accept the flood; (3) They left messages on the infected computers that made up the botnet; and finally (4) issued unspecified commands that disabled the bots on those computers, killing the botnet. There was immediate reaction to the news of this experiment by a law firm in the United Kingdom, citing probably violation of the British Computer Misuse Act by the unauthorized access and use of computer resources, and unauthorized modification of the configuration of the involved computers. The BBC's response to the criticism was to state they had no intention of violating laws, and believed their actions were justified by citing, in their words, "*a powerful public interest in demonstrating the ease with which such malware can be obtained and used; how it can be deployed on thousands of infected PCs without the owners even knowing it is there; and its power to send spam e-mail or attack other Web sites undetected.*" [62]

Lycos Europe: "Make Love not Spam" Campaign. [23] [*Case* 11] In 2004, Lycos Europe – a service company with roughly 40 million e-mail accounts in eight European countries – decided it was time to do something to counter unsolicited commercial email (also known as *spamming*). Lycos created a screen saver designed to impact sites associated with spam emails by consuming the majority of bandwidth available to those sites. The system, and campaign associated with it, was named *Make Love not Spam* (MLNS). The

MLNS campaign began operating in late October 2004, and was ended the first week of December 2004 after the screen saver was installed by over 100,000 users. Their two principle stated goals were punitive and retributive: (1) to annoy spammers and to thereby convince them to stop spamming by (2) increasing their costs and thus decreasing their profits. Lycos did not show they had no other options, such as law suits, by which to achieve the same goals. Lycos could not guarantee specific targeting of only culpable parties, nor did they correlate *illegal* spamming with targeting. Some targets could have been innocent of any criminal acts. The final analysis, based on the principles expressed by Himma, showed Lycos had failed to meet the preconditions of the Defense Principle, the Necessity Principle, or the Evidentiary Principle.

University of Bonn: Stormfucker. [*Case* 12] On December 29, 2008, a research group from the University of Bonn presented a talk at the 25th Chaos Communication Conference (25C3) in Germany on "Owning the Storm botnet." This research was inspired by the Storm enumeration research at the University of Mannheim. [38] The group demonstrated how knowledge gained from reverse engineering the Storm botnet's command and control (C&C) protocol allowed them to take control of Storm nodes. They showed how Storm bots could be commanded to download and replace Storm with *any chosen binary executable.* Such reverse engineering is required for comprehensive understanding of emerging malware threats. [24, 48, 38, 17, 8] Partial source code for their program that implements the counter-attack on the Storm botnet (named *Stormfucker*) was released on the `full-disclosure` mailing list. In their 25C3 presentation, and an interview following the conference, [19] they caution that affecting compromised computers is illegal in many countries, but speculate that someone who resides in a country where there are no laws preventing such action might use the knowledge embodied in the released code to dismantle the Storm botnet, or complete their own working code and publish it. This work was not presented in an academic setting. Had it been, a discussion of the ethical principles that could justify attempting to clean up thousands of infected computers, such as with Denning [22] or Spafford [69], would help guide those with access to the source code in deciding how to use it.

Two of the researchers presented this research at a conference at the Cooperative Cyber Defence Centre of Excellence in Tallinn, Estonia, in June 2009. The abstract of their talk, [52] "asks urgently for political discussions about authorization and legal feasibility" of taking offensive measures to clean computers without their owners' knowledge or consent, and argues that, "pro-actively fighting botnets requires immediate political and international consensus."

Information Warfare Monitor: GhostNet. [20] [*Case* 13] Between June 2008 and March 2009, researchers in Canada conducted a multi-phase investigation of a malicious botnet. The victims included the foreign embassies of dozens of countries, the Tibetan government-in-exile, development banks, media organizations, student organizations, and multinational consulting firms. Initial research involving passive monitoring of suspected victim networks confirmed the intrusions and identified the malware, which was then reverse engineered. *Honeypots* were then infected and used to collect intelligence on the botnet's operation and control servers. The researchers "scouted these servers, revealing a wide-ranging network of compromised computers." Gaining access to the attackers' command and control front end, they were able to, "derive an extensive list of infected systems, and to also monitor the systems operator(s) as the operator(s) specifically instructed target computers." [20] This activity falls within the lower- to mid-level of aggressiveness in the Active Response Continuum, [25] and most certainly involves unauthorized access to systems outside of the authority of the researchers. While there is not mention of ethical considerations, the researchers' actions appear to conform with the ethical issues of proportionality, defense, necessity, and are narrowly targeted at attacker-controlled systems. It is assumed from the structure of the report that it was delivered to law enforcement agencies directly or indirectly through the victims being assisted.

Tipping Point: Kraken botnet takeover. [59] [*Case* 14] In May 2008, researchers at TippingPoint Technologies' Digital Vaccine Laboratories reverse engineered the encryption used by the Kraken bot, and were able to infiltrate and take control of the 400,000 host botnet. This is the same activity performed by some academic research groups, and results in the same situation: the potential to fully control a malicious botnet. One of the researchers interviewed, Cody Pierce, suggests they were, "one click away from [shutting] down the communication between the people sending commands to these [infected] computers." While they may have had no intention of taking action, the discussion surrounding the situation is applicable here. A statement by Endler (tipping point) is interesting to consider: *If you see someone breaking a window to go into someone's house, that really doesn't give you the right to break another window and go in after them.* [59] Implicitly, Endler is talking about violating a third-party's property rights by breaking in to take action (either punitive or retributive) against a criminal. This would not be justifiable, according to Himma, under any of the ethical principles he cites. There is at least one state court decision, however, that aligns with the *Necessity Principle* [1] in suggesting that an emergency private search may be allowable. The reasoning involves allowing a private citizen to break and enter into another's property to *retrieve and protect the stolen goods* of a victim of theft if they are easily destructible or concealable.

Symbiot: Active Defense. [*Case* 15] In March 2004, the Austin, Texas based company *Symbiot, Inc.* announced a product named the *Intelligent Security Infrastructure Management Systems* (iSIMS) platform possessing counter-strike capabilities. [32] Their product was positioned as a means for victims to not only block detected attacks, but to automatically identify "attackers" and direct retaliatory strikes, or even launch preemptive Denial of Service (DoS) attacks to stop attackers. Critics said the system encouraged vigilantism, and noted that true attribution of attackers was not actually being done, only *last-hop* identification, thus targeting of innocents for the counter-strikes was highly likely. The system was also promoted in terms of allowing retributive and punitive actions.

Tracing Anonymous Packets to Their Approximate Source. [12] [*Case* 16] Burch and Cheswick show a method that

uses controlled flooding of a link using the UDP chargen service to achieve a form of IP traceback to the attacker's source, or close enough to it. At a time when DDoS was on the rise, many methods were being explored to tackle the problem. The researchers even dedicate a small section at the end to the ethics of their approach: they admit that their method could be questionable, perhaps even just as bad as the attack they were trying to trace. However, they argue that their intent was the benefit of the Internet community, whereas the intent of the attacker was to harm the community.

3.3 Vulnerability and Disclosure

LxLabs Kloxo / HyperVM. [*Case* 17] LxLabs, a company based in Bangalore, India, markets a web server virtualization system called *HyperVM*, which uses an administration interface named *Kloxo*. One company who uses HyperVM and Kloxo is UK-based Vacert.com. On Sunday, June 7, 2009, Vacert.com suffered a compromise of their web hosting system, resulting in over 100,000 accounts being deleted from the system. On Monday, June 8, 2009, LxLabs' CEO, 32 year old K T Ligesh, was found dead in his apartment of an apparent suicide. [53]

Just a few days before, on June 6, 2009, an analysis of "several dozen vulnerabilities in kloxo" with complete details on how to exploit these vulnerabilities was posted anonymously to the web site *milw0rm* [73] . The time line in this analysis describes an attempt by the unknown security researcher to correspond with staff at LxLabs about the vulnerabilities, which includes such problems as file permission bypass, cross-site scripting, symbolic link exploitation, denial of service, and arbitrary command execution at elevated privilege through unclean user input. The posting claims an initial report was sent two weeks prior (on May 21, 2009) and that resources demonstrating the vulnerabilities were made available to assist LxLabs in confirming and fixing the problems. After two email exchanges with an unnamed LxLabs employee, no further communication as promised from LxLabs, and no observed attempts by LxLabs to even review the resources, the researcher posted the full analysis and exploit details. Within days, multiple sites using Kloxo (including Vacert.com) were attacked by unknown parties. As the researcher points out, the set of vulnerabilities he or she discovered are in stark contrast to the claims made by LxLabs about the security of its product.

It is not known whether there is any relationship between the person(s) who attacked and damaged the web sites and the security researcher who published the vulnerability information, nor the identity of the person who the researcher was in communication with at LxLabs (e.g., the CEO, or some else.) There is no indication that the security researcher attempted to report these problems to any other organizations (such as CERT/CC or other CERTs, news organizations, etc.). Finally, there is no indication that the researcher considered releasing only partial details in order to warn Kloxo users or the general public and give them a chance to protect themselves (perhaps by switching providers or backing up their web sites) prior to release of full details including exploits, as is recommended in various responsible vulnerability disclosure guidelines [60] and policies [4].

Exploiting open functionality in SMS-capable cellular networks. [29] [*Case* 18] Enck et al. suggest a bandwidth-exhausting attack on cellular networks by sending enough text messages (SMSs) to prevent establishment of voice channels for legitimate callers. Since text messages and voice-setup messages use the same medium, this attack is possible, which is what the authors clearly demonstrate in their paper. According to the authors, a sufficiently dedicated attacker can disrupt voice traffic for large cities such as New York, and a truly dedicated attacker can target a large continent with the help of a DDoS network. They provide the required message rate for a successful attack on cities like New York or Washington, DC. They offer some thoughts on how to solve or mitigate this problem, but the solution does not appear to be complete without a complete re-architecture of the cellular network. They suggest that this problem should be investigated further to protect this critical infrastructure.

Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses. [34] [*Case* 19] Implantable cardiovascular defibrillators (ICD) are implanted medical devices used to sense a rapid heartbeat and administer a shock to restore a normal heart rhythm. These devices are configurable through a device programmer which connect to the ICD wirelessly. This paper demonstrates several attacks on the privacy and and integrity of one such medical device using a software programmable radio. The proof of concept attacks described in the paper determined if the patient had such a device, its type, personal information about the patient and reception of real time telemetry data. More importantly, the attacks showed the ability to change or disable therapies (what the device does in certain conditions) and the ability to deliver commands to shock the individuals heart. The potential risks of such a disclosure have immediate and life threatening impacts. As such, they are fairly anomalous when compared with the risks associated with most security research. The authors go to great lengths to avoid discussing of attacks from distances ($\gtrsim$ 1 CM), attack or protocol specifics, or descriptions of how their attacks could impact the health of an individual. The authors intentionally explore the rational for their disclosure, in spite of the risk, describing the benefits in terms of increased privacy and integrity for future such devices.

Black Ops 2008 – Its The End Of The Cache As We Know It. [47] [*Case* 20] In the summer of 2008, Dan Kaminsky (IOActive, Inc.) found a practical attack on an old bug involving a weak random number generation algorithm used for creating transaction IDs. These transaction IDs were meant to ensure clients were talking to the real DNS server. The bug existed in dozens of popular DNS implementations serving between 40% to 70% of internet users.

Attackers exploiting this bug could poison DNS cache entries and control where victims' computers connected. As DNS is critical to operation of all services on the internet, and plays a key role in a wide variety of trust chains, significant damage could result from widespread exploitation of this bug. Balancing the huge risk, the author intentionally set about the process of notification and correction *before* publication/presentation at Blackhat, including the controversial step of requesting that other researchers *not speculate*

on the bug or develop attacks of their own. As a result of patient and coordinated disclosure and mitigation efforts, hundreds of millions of users were protected prior to the vulnerability being announced.

RFID Hacking. [11] [*Case* 21] Bono et al. revealed a hardware mechanism, built from publicly accessible resources, for breaking RFID devices used in the SpeedPass, a payment token for purchasing gasoline and other items at a US gas station, and also in RFID-enabled car ignition keys. Their approach included reverse engineering the device, showing that it was possible to crack the 40-bit key in roughly an hour, and creating a cloned device with which they purchased gasoline, and also starting a car with a cloned device.

Heydt-Benjamin et al. [35] built a device to capture and clone first-generation RFID-enabled credit cards. This earned them a related episode in the then popular US television show '24'. As they show, the credit card owner's data can be captured at a distance, e.g. by pointing a reader at the person or their purse to access the RFID chip. To demonstrate their work, they successfully completed a purchase with their cloned device using a commercial credit card reader.

How to Own the Internet in Your Spare Time. [71] [*Case* 22] Staniford et al. start by analyzing Code Red, comparing it to Nimda, and speculate about future worms by exploring various propagation vectors. They create conceptual worms, such as an improved Code Red (aptly named Code Red II), flash worms, hit-list scanning worms, the Warhol worm, and the topological worm, and muse about their propagation speeds and control vectors. They also explore the concept of a stealthy contagion of users via file-sharing networks. In summary, they provide several recipes for creating massive disruptions within a short period of time.

Botnet design. [*Case* 23] In "Army of botnets" [74] and "An advanced hybrid peer-to-peer botnet' [75], the authors devise botnets based on smaller disjoint botnets that collude to form a much larger botnet, or advanced command and control mechanisms for P2P botnets. In either case, the level of description for the mechanisms is very high, from pseudo-code to the key exchanges necessary to create and maintain such advanced botnets.

WORM vs. WORM: preliminary study of an active counter-attack mechanism. [15] [*Case* 24] Castaneda et al. propose the concept of anti-worms, an automated process that generates a variant of the worm in question. They created a Windows-based prototype and tested it in a smaller run, and simulated its effects at a larger scale. Some of the proposed mechanisms include a patching worm, one that would either remove an existing worm infection or prevent it altogether The authors do realize that there are some legal issues (accessing a remote computer without the consent of the user) and network implications (disruptions by spreading just as fast as the original worm) for their approaches and present a short discussion to that effect. When this paper was published, concepts like Code Green and CRClean, anti-worms for Code Red, had already been publicly discussed.

A pact with the devil. [10] [*Case* 25] Bond and Danezis create the Satan Virus, aka The Devil Worm, a hypothetical ultimate worm that plays the participants against each other. The propagation of the malware is drawn by temptation of access to another user's machine, mails, and documents in general. It further tempts the infected user to recruit more targets for it, since it watches the infected user for remote access to the machines it originally gave the infected user access to. By threatening to disclose this unauthorized access, the malware then blackmails the user to continue gathering new users for its network, and then eventually double-crosses the user and "sells" his or her information as well. While the malware is hypothetical, the authors do describe implementation issues and sample temptations and threats that the malware can use.

3.4 Publication of Results and Data

Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT). [2] [*Case* 26] The availability of realistic network data plays a significant role in fostering collaboration and ensuring U.S. technical leadership in network security research. Unfortunately, a host of technical, legal, policy, and privacy issues limit the ability of operators to produce datasets for information security testing. The Virtual Center for Network and Security Data is a unique effort to organize, structure, and combine the efforts of the network security research community with the efforts of the Internet data measurement and collection community. Under the umbrella of the Protected Repository for the Defense of Infrastructure against Cyber Threats (PREDICT) initiative of the Department of Homeland Security Science and Technology directorate, the Virtual Center will provide a common framework for managing datasets from various Internet data providers. It also will formalize a process for qualified researchers to gain access to these datasets, in order to prototype, test, and improve their Internet threat mitigation techniques, while ensuring that the privacy and confidentiality of Internet users are not compromised. Does publication of network data effect the privacy of Individuals? Can the government sponsor this research? Are current privacy protection methods (anonymization) sufficient? Is it legal for providers to collect this data?

4. SYSTEMATIC EVALUATIONS OF CASE STUDIES

Bynum and Rogerson [14] suggest a multi-staged approach to case study analysis in order to build ethical judgement capabilities. These stages include: identifying key ethical principles, detailing the case study, identifying specific ethical issues raised by the case, calling on your own experience and skills for evaluation, then the abilities of others, and finally, applying a systematic analysis technique.

We detail the case studies and provide an analysis of the ethical issues raised in Section 3. However, in order to make best use of those studies, we first identify the key ethical issues for security researchers and extend existing frameworks so they can be used as a systematic analysis technique.

4.1 General Ethical Issues for Security Researchers

When considering actions related to research or mitiga-

Table 1: Potential Ethics Issues. (• = Central Ethical Issue, ○ = Tangential Ethical Issue)

Principle	Question	Case Number															
		10	12	5	6	7	11	8	15	2	22	1	4	16	24	23	19	26
Defense	Population being protected is identified?	•	•	•	•	•	•	•	•	•	○	○	○	•	○	○	○	○
Defense	Looks like use of *force*?	•	•	○	○	•	•	○	•	○			○	•			○	
Defense	Actions are proportional?	•	•	○	○	•	•	○	•	○			○	•				
Defense	Necessary to repel or prevent harm?	•	•	•	•	•	•	•	•	•	○	•	•	•	○	○	○	○
Defense	Benefits of disclosure favor victims over attackers?	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Defense	Actions are appropriately directed?	•	•	•	•	•	•	•	•	•		○	○	○			○	
Necessity	Greater moral good defined?	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Necessity	No other reasonable options available?	•	•	•	•	•	•	•	•	•	○	○	○	•	○	○	•	○
Necessity	Otherwise respectful of rights?	•	•	•	•	•	•	•	•	•		○	○	•			•	○
Punishment	Avoids punitive motives?	○	○				•		•					○				
Retribution	Avoids retributive motives?	○	○				•		•					○				
Evidentiary	Adequate reason to think preconditions of applying other principles are met?	•	•	•	•	•	•	•	•	•	○	•	•	•	•	•	•	•
Respect for Persons	Individuals treated as autonomous agents?	•	•		•	•	•		•	•		•	•	•			○	
Respect for Persons	Individuals (or their providers) informed and allowed to consent?	•	•	○	•	•	•	○	•	•		•	•	•			•	•
Respect for Persons	Individuals with diminished autonomy are protected?	•	•	○	•	•	•	○	•	•		○	•	•			•	•
Respect for Persons	Identities of innocents are protected?	•		•	•	•	•	•	•	•		•	•	•			•	•
Beneficence	Low potential to inflict harm?	•	•	○	•	•	•	○	•	•	○	•	•	•	○	○	•	•
Beneficence	Maximize possible benefits and minimize possible harms?	•	•	○	•	•	•	○	•	•	○	•	•	•	○	○	•	•
Beneficence	Risks and benefits systematically assessed?	•	•	○	•	•	•	○	•	•	○	•	•	•	○	○	•	•
Justice	Who benefits?	•	•	○	•	•	•	○	•	•	○	•	•	•	○	○	•	○
Justice	Fairness (neutrality) of procedures & outcomes in selection of subjects?	•	•	○	○	•	•	○	•	•		•	•	•			•	○

tion of malicious or illegal activity, there are many issues that must be considered. These involve issues of (a) proportionality, (b) targeting, (c) necessity, (d) desired outcome, (e) potential consequences, and (f) the greater moral good to society that is expected to result (and whether it outweighs any potential harm to innocent third parties.)

For example, there are situations in which great tension exists between releasing information to claim first discovery, or holding it back to prevent harm. This is uncommon to most academic research, where discoveries are primarily applied immediately to benefit society. We must remember that much computer security research is focused on *criminal activity* that is actively causing harm to the public, and the potential for harm from unethical actions could extend to the entire internet population. Take public/private partnerships aimed at responding to cyber threats, which involve government, commercial entities, academic researchers, and select members of the public who specialize in computer crime activity. Here the tension to go public with new knowledge is more intense. Vendors of anti-malware products may wish to be first to disclose to increase their customer base and sell more products. Academics may wish to be first to disclose to enhance their academic positions and increase chances of future funding. Private citizens may wish to improve their chances of getting a new high-paying job. In research into curative treatments in healthcare, premature disclosure of study results will not be used by viruses, microorganisms, or cancer, to improve pathological efficiency (but pharma certainly cares!) In security research, such premature disclosure *can and does* result in improvements of malicious software and tactics that make the task of responders much harder, and the potential harm to the public greater.

The kinds of questions that researchers must ask themselves include (but are not limited to) the following:

- Are the research results intended to protect a specific population, and if so, which population? (E.g., the owners of infected hosts, the victims of secondary attacks using a botnet, the researchers' own institution, or the general internet user?)
- Is there a way to achieve multiple benefits to society simultaneously when studying criminal botnet behavior? (E.g., developing new defenses, while aiding investigation of criminal acts and assisting victimized network sites?)
- Who will benefit more from publication of research findings, and in which order: Victims of criminal acts; authorities responsible for protecting their citizens; the researchers themselves; or the criminals who are perpetrating computer crimes?
- Is there any other way to accomplish the desired research result(s)?
- What is the safest way to disseminate research results without risk of improper use by individuals who may not share the researchers' ethical standards?
- If all security research is halted because 100% safety cannot be guaranteed, is the result a greater *harm* to society because no new defenses are developed, or is taking the risk of some small number of potential infections worth the thousands or millions of hosts protected by resulting new defenses?

While these general questions get to some of the issues, they are not sufficient to give fine-grained guidance in a form that could be evaluated. We are encouraging researchers to include in publications an indication that they have made

the effort to evaluate their work against the ethical questions raised in this paper in a way that is uniform across all research situations and topics. Further, using a simple and uniform methodology supports consistent evaluation by outside parties in a manner that improves trust in computer security research protocols.

4.2 Towards a Systematic Approach

Table 1 shows a potential ethics scoring guide that includes the salient ethical principals and their sub-components as identified in the previous Sections. The table is split into two sets. The issues on top of the table come from the Active Response Continuum, and are primarily aimed at situations involving direct interaction with hosts outside one's own administrative control. The issues on the bottom come from human subjects regulations (Section 2.4.3) and are primarily aimed at protection of research subjects (and for the purposes of this paper, other indirectly involved third-parties.)

For example, the Punitive and Retributive Principles only apply in situations where there is active manipulation in some form of external systems that is directed at *attackers* in any way. This is also true of aggressive actions that may not be punitive or retributive, but still may, "look like use of force." These actions could only be justifiable if properly directed at the culpable party (the attacker), and not systems owned and operated by innocent third parties, and also be proportional. It is doubtful these would ever fit into a research protocol, but those who may extend work from the research community might try to cross that line.

A set of 16 representative case studies were chosen from Section 3 and evaluated as to the ethical issues raised. Filled circles indicate the issue is *central* to the case study, and empty circles indicate the issue is *tangential*, or of lesser importance. Those cases involving active engagement with third party systems in some way (e.g., internal botnet enumeration, disinfecting, monitoring, taking control of botnet command and control, copying files, etc.) tend to involve the most issues, while those that are more narrowly focused (e.g., on vulnerability disclosure) tend to involve fewer.

Following Bynum and Rogerson, we have identified how these issues pertain to the selected cases. We do not take the next step and perform an evaluation, however this could follow in future work. The work we have done to date provides a straw-man proposal for a foundation on which a scoring methodology could be built, and a simple and clear set of issue laid out to guide researchers in developing their research protocols in a unified manner. The result can establish a basis of trust with the general public, who are implicitly involved as central stake holders.

Our aim intent is to find a way for researchers to take risks that are acceptable to the general public and address the advancing threat landscape. Being overly conservative may lose ground to these threats. Being overly aggressive or risky, especially if some harm results, may generate a backlash that likewise loses ground. A reasoned and measured approach, based on accepted ethical standards, can result in a decreased threat landscape. It can also result in something akin to the legal concept of a *reasonable person* standard (i.e., "Would a reasonable person, in the same circumstances, chose to act in the same way?")

4.2.1 Stakeholder Analysis

Stakeholder Analysis identifies the key players in the situation in terms of their interests, involvement, and their relationship (i.e., producer or recipient) of outcomes such as benefit or harm. We will adapt the definitions of stakeholders from the Association of Social Anthropologists of the UK and Commonwealth for the purposes of this section [54].

Primary stakeholders are, "those ultimately affected, either [positively or negatively]." These will typically be the end-users of computer systems, and consumers of information or information system products or services.

Secondary stakeholders are, "intermediaries in delivery" of the benefits and harms. In the computer security context, these would be service providers, operators, or other parties responsible for integrity, availability, and confidentiality of information and information systems.

Key stakeholders are, "those who can significantly influence, or are important to the success [or failure] of the project." We will include the researcher(s), vendor(s), those who design and implement systems, and criminals or attackers. We include the latter because malicious activity is a direct cause of action, just as much as a manager ordering an engineer to make a design change or fix a software bug.

4.2.2 Roles and Responsibilities Analysis

Roles and Responsibilities Analysis takes the identified Stakeholders, and lists both their role or roles in the situation, as well as their responsibilities towards each other and to society as a whole.

Once stakeholders have been identified, and roles and responsibilities mapped out, one can start to define desired outcomes in terms of maximizing benefits and minimizing harms to stakeholders. Alternative actions that fall within the delineated roles and responsibilities can then be weighed against each other in terms of expected outcomes. One of the hardest challenges is in trying to identify potential negative outcomes that may result from alternative actions in order to minimize unintended consequences. This is where involvement of trusted external parties, such as peer-review of proposed actions or protocols, can help.

5. RESEARCH CHECKLIST

- Does your research involve human subjects? Can it successfully be replaced by random processes? [31]

6. RELATED WORK

There have been various works that focus on ethical issues in the information security context over the years. Each takes a different, and sometimes very narrow, look at the subject.

Denning (1990) and Spafford (1992) discuss ethics in the context of those engaged in *computer intrusions*. Denning [21] describes the opinions of hackers who were interviewed about computer intrusions as to whether those acts were ethical or not. Some hackers believed that certain malicious actions where wrong and unethical (e.g., "breaking into hospital systems," "reading confidential information about individuals," "stealing classified information," "committing

fraud for personal profit.") Some hackers believed that exploring computer systems was ethical, provided that "the objective is to learn and avoid causing damage." Spafford [69] looks at similar acts in terms of *right and wrong*, and whether a *greater good to society* is achievable by computer intrusions. In Spafford's analysis, computer intrusions may only be ethically justifiable in the most extreme cases, such as to save a human life in an emergency. In discussing publication of worm or virus code (which may be capable of resulting in harm to innocent third-parties) he states that, "publication should serve a useful purpose; endangering the security of other people's machines or attempting to force them into making changes they are unable to make or afford is not ethical."

Greater attention has been paid to the ethics of *responding* to computer attack in terms of *use of force* alternatives under theories of the *Law of War*, or *Law of Armed Conflict* (LOAC). For the purposes of this discussion, there are parallels to concepts embodied in the LOAC. For example, the LOAC requires *military necessity* as a pre-requisite for the use of force. It requires *distinction*, that is, actions must be directed against lawful combatants and military targets, not against civilians and civilian infrastructure. Lastly, the LOAC requires *proportionality*, that is a use of force less than or equal to the original harm or violation. As a result of international agreements and protocols, such as those defined in the Geneva Conventions, [39] militaries around the world operate under strict *Rules of Engagement* (ROE). These ROE guide decision making on the field of battle to ensure the actions of military personnel do not result in potential war crimes charges.

Yurcik (1997, 2000) [78, 79] discusses ethics in relation to attack and retaliation using *information warfare* (IW) tactics, and considers whether the lethality of IW operations affects the ethics of employing such operations in defense. This applies to military responses to attacks at the nation-state level, but sets the stage for the equivalent considerations of responses in non-military settings. Yurcik next considers *hack-back*, or aggressive responses to computer attack by attacking back. [44] More complete analyses of the application of international law and the law of war to state-directed IW – also known as *cyberwarfare* – operations were done by Sharp (1999) [70] and Wingfield (2000). [76] The actors here are primarily nation-states, not individuals.

Dittrich and Himma (2004) [25] discuss the legal and ethical frameworks for responding to computer intrusions. This work follows from a workshop and research led by Dittrich in 2003, which were inspired by an initial workshop on *Active Defense* organized by Kirk Bailey of Seattle's *Agora* security group in 2001. Their research identifies three ethical principles as being central to consideration of aggressive counter-measures: the *Defense Principle*, the *Necessity Principle*, and the *Evidentiary Principle*. Dittrich and Himma build on previous work by Yurcik, specifically focusing on the non-military considerations for response, as well as considering transition of response from civilian to military realms. Himma later expands [36] on previous work with Dittrich to include the *Punishment Principle* and the *Retaliation Principle*.

Denning (2007) [22] evaluates the traceback activities of Sean Carpenter related to the reported computer incursions collectively known as *Titan Rain*. Using a *Schmitt analysis*, [56] Denning concludes that Carpenter's traceback does not look like force, and thus could potentially be justifiable. Denning also analyses a more aggressive action, such as a hack-back to clean up compromised computers (bots) involved in a DDoS attack. This, she argues, may not be justifiable as proportionate, has no affect on eliminating the original source of the attack, and has a serious risk of causing greater harm to the systems involved than their original participation in the botnet. Throughout her analyses, Denning brings up many points that highlight the complexity of the ethical choices surrounding alternative actions.

Burstein (2008) [13] discusses collecting network traces, running infected hosts, publishing data, and mitigation. The primary focus is on law, though he raises three basic ethical questions: (1) whether research activity would harm the reputation of the researcher's institution, (2) whether publication would aid an adversary in retaliating, and (3) whether it is a researcher's role to engage in botnet mitigation.

7. FRAMEWORK APPLICATION

In this section we will use some of the ethical frameworks and analysis techniques described in previous sections to examine a few selected case studies from Section 3. This will allow us to apply our proposed framework and illustrate some of the alternatives and how to decide which choices will achieve the best outcomes.

7.1 Example: LxLabs Kloxo / HyperVM

The study of *Case* 17 is interesting and unique in terms of the possible relationship with the suicide of the CEO of the vendor. It is not a situation in which academic research is involved, but it does bring in many issues of risk/benefit, disclosure of vulnerability information, and consideration of alternative courses of action. In *Case* 17 , we can identify the following stakeholders:

- The *researcher* who discovered the vulnerabilities. (This person has chosen to remain anonymous.) [Key stakeholder]
- The *programmers* who were responsible for creating the HyperVM system and Kloxo administrative front end. [Key stakeholder]
- The *corporate management of the vendor* (LxLabs), which includes the *CEO*. [Key stakeholder]
- The *service providers* who purchased HyperVM / Kloxo. [Secondary stakeholder]
- *Criminals and Attackers* who would exploit vulnerabilities for their own purposes. [Key stakeholders]
- The *customers* of the service providers who use the virtual servers. [Primary stakeholder]
- The *consumers* who obtain products or services from the customers of the service providers (e.g., the online merchants using virtual storefronts hosted on HyperVM virutal machines.) [Primary stakeholder]

The researcher attempted to contact the corporate management of LxLabs, presumably to convince them to make decisions that would direct the programmers to fix the bugs that the researcher identified. Implicitly, we assume the researcher chose to contact the vendor privately to allow

them to fix the problem in order to protect the primary stakeholders (i.e., virtual machine customers and their end consumers.)

The action of the researcher as a key stakeholder to make detailed vulnerability and exploit information to the vendor is intended to assist the vendor in correcting the problems and eliminating the vulnerability. This creates a benefit to the primary stakeholders by protecting their services and accounts, as well as benefiting the secondary stakeholders by improving their product and protecting their customers.

It is the vendor's responsibility as a key stakeholder to use this information to minimize potential harm to the primary stakeholders. While the researcher did not state this explicitly, we can assume that the researcher has taken upon himself/herself the responsibility of assisting in protecting the primary and secondary stakeholders. We can infer that the action of reporting was intended to obtain the outcome of protecting the primary stakeholders by minimizing harm to them that would result from a malicious actor finding and exploiting these vulnerabilities before the vendor corrected them.

The researcher had several alternative pathways that could achieve this same goal:

- The researcher could have taken a high-level outline of the vulnerabilities and provided them to a reporter, who could have written a news article disclosing (in general terms) that vulnerabilities in the HyperVM / Kloxo system were discovered and warning the primary stakeholders (i.e., customers and end consumers). The primary stakeholders could then take their own actions to ask questions, harden defenses, ensure they had current backups, or consider moving their storefronts to other service providers. These are all results that minimize harm.
- The researcher could have identified a representative set of HyperVM / Kloxo customers and warn them (again, in general terms) of the vulnerabilities and/or provided mitigation details. These customers could have been encouraged to contact LxLabs and put pressure on the vendor to fix the problems. This pressure could come in the form of complaints, threats to find alternative vendors, or threats of lawsuits in the event that these vulnerabilities are exploited and harm results prior to patches being made available. This would also have the same added benefits in terms of mimization of harm as the previous option. This would not be as easy as contacting a single reporter, or reporting to a CERT organization, but would still move towards acheiving the goal of protecting the customers and end consumers.
- The researcher could have published a high-level summary of the vulnerabilities, rather than full exploit details. This may well result in calls from full-disclosure advocates to provide more details, and possibly criticism of the researcher for over-stating the significance of their findings. Anyone with the same (or greater) skills would be able to repeat the research and thus possess the same ability to exploit these vulnerable systems, however this would take time that the vendor may be able to use to fix the problems before any harm is done to the primary stakeholders. The researcher thus has to balance personal benefit from first discovery and/or immediate full disclosure, potential reputational harm resulting from criticism for partial disclosure, and potential harm to primary stakeholders from release of exploit information prior to patches being available to fix the bugs.

The use of anonymity by the reseacher for unstated reasons leaves open many questions. (i) It may indicate that there is no personal gain to the researcher from disclosure. Then again, it also has the potential of avoiding accountability for any actions that are taken, including unintended consequences that cause harm. (ii) Releasing full details only two weeks after first contacting the vendor is another difficult issue. Because there was no evidence of the vendor even looking at the vulnerability details, could the researcher have been acting with a punitive motive against the vendor? If so, this would violate one of Himma's principles. (iii) The researcher may be a disgrunteled current/former employee with a retributive motive. (iv) The researcher may have chosen to use anonymity to protect themselves from civil litigation brought by the vendor for claimed violation of anti-circumvention, computer access, or other computer crime related laws. [27]

The complete exploit details do not help the primary stakeholders in protecting themselves, as there is nothing they can do (short of immediately switching to another virutal machine provider, which would take significant effort and time.) Full details can reasonably be seen to benefit attackers more.

It is unreasonable for the researcher to anticipate the CEO would commit suicide, nor is it provable that the pressure from disclosure and resulting damage from exploitation of the vulnerabilities contributed to the suicide. It is foreseable, however, that disclosure of full exploit details without warning would likely result in one or more parties using this information to do anything made possible, up to and including destroying the contents of any servers they could find. Thus, if the primary goal of the researcher was to minimize harm to the primary stakeholders, the choice to disclose the vulnerabilities two weeks after first attempting to contact the vendor resulted in the exact opposite result. That is, it increased harm and decreased benefit to both primary and secondary stakeholders.

8. FUTURE WORK

We have reviewed many security research situations where those involved faced questions about what they should and should not do with knowledge they possess. In some cases, actions taken were questioned by observers. In other cases, actions were not taken and we will never know if a greater good to society would have resulted, or if any damage to property, lives, or reputations would result. We have also seen frustration expressed by those witnessing the growth in computer crime and desiring something be done about it, and growing interest by researchers and defenders to respond to do just that. But there is insufficient guidance today for researchers to follow, or standards by which to judge research activities. There is even a question of whether academic or private researchers should actively be involved in computer crime activities solely for research purposes (as opposed to supporting protective or investigative activities.) [13]

More questions are typically raised about the ethics of

computer security research activities than are answers provided, which can illuminate topics for future work. What constitutes risk, and who is placed in harms way? Do some research activities themselves come sufficiently close to a *use of force* that they warrant special consideration? Is federal regulation of research of computer crime activity necessary, similar to research into biological agents or toxins like anthrax, ricin, and smallpox (Public Law 107-188)? Or is a government-mandated ethical review model that extends beyond the purview of IRBs, such as the Embryonic Stem Cell Research Oversight (ESCRO) Committees more appropriate and easier to implement [18]? Could the Information Assurance concepts of *integrity, availability, and confidentially* be used to untangle a complex mix of many inter-related actions when making risk/benefit evaluations? Do we need to separate risks to the systems owned by innocent third-parties and their data from the risks to yet other victims whose data may be stored by attackers on the former parties' systems? (That is to say, is there a need for some kind of recursive calculation of risk in the context of complex computer crime situations, rather than a more simplistic bi-directional researcher⇔subject relationship as is typical in biomedical or behavioral research?) How might a scoring system be developed for uniformly evaluating risk, benefit, and appropriate actions?

To help understand these issues better and define a workable ethical framework, we believe that a more structured series of public discussions is urgently needed. We look forward to seeing these discussions accompanying future botnet research.

Acknowledgments

This work is supported in part by the Cisco Systems Critical Infrastructure Assurance Group (CIAG), the U.S. Department of Homeland Security Science & Technology Directorate under Contracts No. NBCHC060090, NBCHC080037, and AFRL cooperative agreement FA8750-08-2-0147 and by NSF under contracts CNS 0627445 and CNS 0831174. We wish to thank Fabian Monrose, Jose Nazario, Christopher Kruegel, John Heideman, Dorothy Denning, Kaiti Carpenter, Michael Collins, Rachna Dhamija, Tanya Matthews, Mike Johnson, and the anonymous reviewers of an earlier version of this text for their valuable suggestions.

9. REFERENCES

[1] People v. Williams, 53 Misc. 2d 1086, 1090, 281 N.Y.S.2d 251, 256 (Syracuse City Ct. 1967).

[2] Protected repository for the defense of infrastructure against cyber threats (PREDICT). `http://www.predict.org`.

[3] Standards And Guidelines. `http://www.shadowserver.org/wiki/pmwiki.php/Shadowserver/StandardsAndGuidelines`, September 2006.

[4] Cert/cc vulnerability disclosure policy. `http://www.cert.org/kb/vul_disclosure.html`, May 2008.

[5] ACM Council. Code of Ethics and Professional Conduct, Oct. 1992. `http://www.acm.org/about/code-of-ethics`.

[6] M. Allman. What ought a program committee to do? In *WOWCS'08: Proceedings of the USENIX Workshop on Organizing Workshops, Conferences, and Symposia for Computer Systems*, pages 1–5, 2008.

[7] S. Baase. *A gift of fire: social, legal, and ethical issues in computing*. Pearson Prentice Hall, 3rd edition, 2009.

[8] M. Bailey, J. Oberheide, J. Andersen, Z. M. Mao, F. Jahanian, and J. Nazario. Automated classification and analysis of internet malware. In *Proceedings of the 10th International Symposium on Recent Advances in Intrusion Detection (RAID'07)*, September 2007.

[9] I. A. Board. RFC 1087: Ethics and the internet. `http://tools.ietf.org/html/rfc1087`, 1989.

[10] M. Bond and G. Danezis. A pact with the devil. In *NSPW '06: Proceedings of the 2006 Workshop on New Security Paradigms*, pages 77–82.

[11] S. C. Bono, M. Green, A. Stubblefield, A. Juels, A. D. Rubin, and M. Szydlo. Security analysis of a cryptographically-enabled RFID device. In *Proceedings of the 14th USENIX Security Symposium*, pages 1–16, 2005.

[12] H. Burch and B. Cheswick. Tracing anonymous packets to their approximate source. In *LISA '00: Proceedings of the 14th USENIX conference on System Administration*, pages 319–328, 2000.

[13] A. J. Burstein. Conducting cybersecurity research legally and ethically. In *LEET'08: Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*, pages 1–8, 2008.

[14] T. W. Bynum and S. Rogerson. *Computer Ethics and Professional Responsibility: Introductory Text and Readings*. Blackwell Publishers, Inc., Cambridge, MA, USA, 2003.

[15] F. Castaneda, E. C. Sezer, and J. Xu. WORM vs. WORM: preliminary study of an active counter-attack mechanism. In *WORM '04: Proceedings of the 2004 ACM Workshop on Rapid Malcode*, pages 83–93, 2004.

[16] V. G. Cerf. RFC 1262: Guidelines for internet measurement activities. `http://tools.ietf.org/html/rfc1262`, 1991.

[17] K. Chiang and L. Lloyd. A case study of the rustock rootkit and spam bot. In *HotBots'07: Proceedings of the First USENIX Workshop on Hot Topics in Understanding Botnets*, 2007.

[18] N. R. C. Committee on Guidelines for Human Embryonic Stem Cell Research. *Guidelines for Human Embryonic Stem Cell Research*. The National Academies Press, 2005.

[19] D. Danchev. Legal concerns stop researchers from disrupting the storm worm botnet, Jan. 2009. `http://blogs.zdnet.com/security/?p=2397`.

[20] R. Deibert, A. Manchanda, R. Rohozinski, N. Villeneuve, and G. Walton. Tracking GhostNet: Investigating a cyber espionage network, March 2009. `http://www.scribd.com/doc/13731776/Tracking-GhostNet-Investigating-a-Cyber-Espionage-Netwo`

[21] D. E. Denning. Concerning hackers who break into computer systems. In *Proceedings of the 13th National Computer Security Conference*, pages 653–664, 1990.

[22] D. E. Denning. The ethics of cyber conflict, June 2008. Chapter 17 in The Handbook of Information and Computer Ethics.

[23] D. Dittrich. How bad an idea was 'Make Love Not

Spam?' Let me count the ways, Mar. 2005. http://staff.washington.edu/dittrich/arc/workshop/lycos-response-v3.txt.

[24] D. Dittrich and S. Dietrich. P2P as botnet command and control: a deeper insight. In *Proceedings of the 3rd International Conference On Malicious and Unwanted Software (Malware 2008)*, pages 46–63, Oct. 2008.

[25] D. Dittrich and K. E. Himma. Active Response to Computer Intrusions. Chapter 182 in Vol. III, Handbook of Information Security, 2005. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=790585.

[26] N. M. Duc. Korea and us ddos attacks: The attacking source located in united kingdom. http://blog.bkis.com/?p=718, July 2009.

[27] Electronic Frontier Foundation. A "Grey Hat" Guide. http://www.eff.org/issues/coders/grey-hat-guide.

[28] Electronic Frontier Foundation. The Coder's Rights Project. http://www.eff.org/issues/coders.

[29] W. Enck, P. Traynor, P. McDaniel, and T. La Porta. Exploiting open functionality in sms-capable cellular networks. In *CCS '05: Proceedings of the 12th ACM conference on Computer and communications security*, pages 393–404, 2005.

[30] P. Finn and M. Jakobsson. Designing ethical phishing experiments. *Technology and Society Magazine, IEEE*, 26(1):46–58, Spring 2007.

[31] S. L. Garfinkel. IRBs and security research: Myths, facts and mission creep. In *Proceedings of UPSEC '08 (Usability, Psychology and Security)*, Apr. 2008.

[32] S. Gaudin. Plan to counterattack hackers draws more fire. http://www.internetnews.com/article.php/3335811, April 2004.

[33] D. Gotterbarn, K. Miller, and S. Rogerson. Software engineering code of ethics. *CACM*, 40(11):110–118, Nov. 1997.

[34] D. Halperin, T. Heydt-Benjamin, B. Ransford, S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. Maisel. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *IEEE Symposium on Security and Privacy*, pages 129–142, May 2008.

[35] T. S. Heydt-Benjamin, D. V. Bailey, K. Fu, A. Juels, and T. O'Hare. Vulnerabilities in First-Generation RFID-enabled Credit Cards. In *Proceedings of the 11th International Conference on Financial Cryptography and Data Security*, February 2007.

[36] K. E. Himma. *Internet Security: Hacking, Counterhacking, and Society*. Jones & Bartlett Publishers, 2007.

[37] T. Holz, M. Engelberth, and F. Freiling. Learning more about the underground economy : A case-study of keyloggers and dropzones. Technical Report TR-2008-006, Department for Mathematics and Computer Science, University of Mannheim, December 2008.

[38] T. Holz, M. Steiner, F. Dahl, E. W. Biersack, and F. Freiling. Measurements and mitigation of peer-to-peer-based botnets: a case study on storm worm. In *LEET'08: First USENIX Workshop on Large-Scale Exploits and Emergent Threats*, Apr. 2008.

[39] ICRC. The Geneva Conventions: the core of international humanitarian law, January 2006. http://www.icrc.org/Web/Eng/siteeng0.nsf/htmlall/genevaconventions.

[40] IEEE Board of Directors. IEEE Code of Ethics, Feb. 2006. http://www.ieee.org/portal/pages/iportals/aboutus/ethics/code.html.

[41] Institute for the Management of Information Systems. Code of Ethics. http://www.imis.org.uk/about/codeofethics/code_ethics.pdf.

[42] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[43] M. Jakobsson, N. Johnson, and P. Finn. Why and how to perform fraud experiments. *IEEE Security and Privacy*, 6(2):66–68, 2008.

[44] V. Jayaswal, W. Yurcik, and D. Doss. Internet hack back: counter attacks as self-defense or vigilantism? In *2002 International Symposium on Technology and Society*, pages 380–386, 2002.

[45] J. P. John, A. Moshchuk, S. D. Gribble, and A. Krishnamurthy. Studying Spamming Botnets Using Botlab. To appear in Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation (NSDI '09), Apr. 2009.

[46] D. G. Johnson and K. W. Miller, editors. *Computers Ethics*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009.

[47] D. Kaminsky. Black Ops 2008 – It's The End Of The Cache As We Know It. In *Black Hat Briefings USA 08*, Las Vegas, Nevada, USA, July 2008.

[48] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: an empirical analysis of spam marketing conversion. In *CCS '08: Proceedings of the 15th ACM conference on Computer and communications security*, pages 3–14, 2008.

[49] C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, and S. Savage. The Heisenbot Uncertainty Problem: Challenges in Separating Bots from Chaff. In *LEET'08: First USENIX Workshop on Large-Scale Exploits and Emergent Threats*, April 2008.

[50] kc claffy. Ten things lawyers should know about the internet. CAIDA Web site, 2008. http://www.caida.org/publications/papers/2008/lawyers_top_ten/.

[51] S. Kelly. BBC team exposes cyber crime risk, Mar. 2009. http://news.bbc.co.uk/2/hi/programmes/click_online/7932816.stm.

[52] F. Leder, T. Werner, and P. Martini. Proactive Botnet Countermeasures – An Offensive Approach. In *Cooperative Cyber Defence Centre of Excellence Tallinn, Estonia*, March 2009.

[53] J. Leyden. LxLabs boss found hanged after vuln wipes websites. http://www.theregister.co.uk/2009/06/09/lxlabs_funder_death/, June 2009.

[54] S. Mascarenhas-Keyes. Ethical dilemmas in professional practice in anthropology. http://www.theasa.org/networks/apply/ethics/analysis/stakeholder.htm, July 2008.

[55] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and

D. Sicker. Shining light in dark places: Understanding the Tor network. In *The 8th Privacy Enhancing Technologies Symposium*, pages 63–76, 2008.

[56] J. B. Michael, T. C. Wingfield, and D. Wijesekera. Measured responses to cyber attacks using schmitt analysis: A case study of attack scenarios for a software-intensive system. In *COMPSAC '03: Proceedings of the 27th Annual International Conference on Computer Software and Applications*, page 621. IEEE Computer Society, 2003.

[57] J. H. Moor. What is computer ethics? *Metaphilosophy*, 16(4):266–275, 1985.

[58] T. Mullen. The right to defend. `http://www.securityfocus.com/columnists/98`, July 2002.

[59] R. Naraine. Kraken botnet infiltration triggers ethics debate, May 2008. `http://www.eweek.com/c/a/Security/Kraken-Botnet-Infiltration-Triggers-Ethics-Debate/`.

[60] National Infrastructure Advisory Council. `http://www.dhs.gov/xlibrary/assets/vdwgreport.pdf`, January 2003.

[61] H. Phong. Korean agency accuses BKIS of violating local and int'l law. `http://english.vietnamnet.vn/reports/2009/07/859068/`, July 2007.

[62] B. Prince. BBC responds to botnet controversy, Mar. 2009. `http://www.eweek.com/c/a/Security/BBC-Responds-to-Botnet-Controversy/`.

[63] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis. My Botnet Is Bigger Than Yours (Maybe, Better Than Yours): Why Size Estimates Remain Challenging. April 2007.

[64] W. Rehnquist. Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), June 1993. `http://laws.findlaw.com/us/509/579.html`.

[65] Several. Results of the Distributed-Systems Intruder Tools Workshop. CERT/CC, Dec. 1999.

[66] D. C. Sicker, P. Ohm, and D. Grunwald. Legal issues surrounding monitoring during network research. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 141–148, 2007.

[67] B. Snow and C. Brooks. Privacy and security: An ethics code for u.s. intelligence officers. *Commun. ACM*, 52(8):30–32, 2009.

[68] T. Son. BKIS plans to sue network security agency for defamation. `http://www.thanhniennews.com/society/?catid=3&newsid=51281`, July 2007.

[69] E. H. Spafford. Are computer hacker break-ins ethical. In *Deborah G Johnson and Helen Nissenbaum, editors, Computers, Ethics & Social Values*, pages 125–135. Oxford University Press, 1992.

[70] W. G. S. Sr. *CyberSpace and the Use of Force*. Aegis Research Corporation, 1999.

[71] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium*, pages 149–170, Aug. 5–9 2002.

[72] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna. Your Botnet is My Botnet: Analysis of a Botnet Takeover. Technical report, University of California, May 2009.

[73] Unknown. Kloxo 5.75 (24 issues) multiple remote vulnerabilities. `http://www.milw0rm.com/exploits/8880`, May 2009.

[74] R. Vogt, J. Aycock, and M. J. J. Jr. Army of botnets. In *Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007)*, pages 111–123, February 2007.

[75] P. Wang, S. Sparks, and C. C. Zou. An advanced hybrid peer-to-peer botnet. In *HotBots'07: Proceedings of the First USENIX Workshop on Hot Topics in Understanding Botnets*, 2007.

[76] T. C. Wingfield. *The Law of Information Conflict*. Aegis Research Corporation, 2000.

[77] N. Wyler. *Aggressive Network Self Defense*. Syngress, 2005.

[78] W. Yurcik. Information warfare: Legal & ethical challenges of the next global battleground. In *Proceedings of the 2nd Annual Ethics and Technology Conference (Ethics'97)*. Loyola University Chicago, June 1997.

[79] W. Yurcik. Information warfare survivability: Is the best defense a good offense? In *Proceedings of the 5th Annual Ethics and Technology Conference*. Loyola University Chicago, July 2000.

View publication stats