Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Department of Homeland Security Menlo Report

January 3, 2012

Authors

This Companion document was inspired by discussions in the Menlo Report working group meetings held over a period of sixteen months. The authors of this Companion document and the Menlo Working Group participants are listed below.

- Michael Bailey, University of Michigan
- Aaron Burstein, University of California Berkeley
- KC Claffy, CAIDA, University of California San Diego
- Shari Clayman, DHS Science & Technology
- David Dittrich, University of Washington, *Co-Lead Author*
- John Heidemann, University of California, ISI
- Erin Kenneally, CAIDA, University of California San Diego, *Co-Lead Author*
- Douglas Maughan, DHS Science & Technology
- Jenny McNeill, SRI International
- Peter Neumann, SRI International
- Charlotte Scheper, RTI International
- Lee Tien, Electronic Frontier Foundation
- Christos Papadopoulos, Colorado State University
- Wendy Visscher, RTI International
- Jody Westby, Global Cyber Risk, LLC

Contents

A Introduction

Researchers are faced with time-driven competitive pressures to research and publish, to achieve tenure, and to deliver on grant funding proposals. That ethical considerations can be incongruent with these incentives is neither novel nor unique to *information and communication technology* (ICT) [55] research. Those conducting ICT research (ICTR) do, however, face a different breed of tensions that can impact research ethics risks. Unfortunately, institutionalized guidance on the protection of research subjects has not kept pace with the rapid transformations in information technology and infrastructure that have catalyzed changes in research substance and mechanics.

The *Menlo Report* [17] summarizes a set of basic principles to guide the identification and resolution of ethical issues in research about or involving ICT. It illuminates a need to interpret and extend traditional ethical principles to enable ICT researchers and oversight entities to appropriately and consistently assess and render ethically defensible research. The framework it proposes can support current and potential institutional mechanisms that are well served to implement and enforce these principles, such as a research ethics board (REB).

This document is a living complement to the *Menlo Report* that details the principles and applications more granularly and illustrates their implementation in real and synthetic case studies.

A.1 Historical Basis for Human Subjects Research Protections

One of the watershed events that focused widespread attention on research ethics was the Nuremberg Doctors Trial following World War II which revealed the forced medical experimentation on Nazi-held prisoners of war throughout Europe. It motivated the 1947 Nuremberg Code which was unprecedented in its call for informed consent and voluntary participation in research experiments. Seven years later, the renowned Declaration of Helsinki was initially drafted by the Medical Ethics Committee of the World Medical Association and later adopted in 1964. It addresses issues with research protocols involving humans in terms of *risks and benefits*, *informed consent*, and *qualifications of researchers*, and informed a set of international standards that apply to clinical research known as Good Clinical Practices (GCP). There are now over one thousand laws, regulations, and guidelines worldwide that protect human subjects of research [45].

In the United States, one of the most infamous biomedical research abuse cases involved experiments on low income African-American males in Tuskegee, Alabama. Commencing in 1932, these studies continued for decades until revealed to the public in 1972. Subjects were purposefully infected with syphilis, or penicillin was not administered to men diagnosed with syphilis despite the widespread knowledge obtained in the 1940s that the drug was an effective treatment. Doctors wanted to observe the longitudinal effects of the disease on patients up until their deaths. In 2010, the United States Government apologized for sponsoring similar experiments on Guatemalan citizens – with the cooperation of health officials in Guatemala – from 1946 to 1948. Lesser known research abuses, such as injecting cancer cells into live patients and performing experimental surgical procedures on patients without their knowledge or consent (often injurious to the point of death) have made their way into the press and the courts [52].

The U.S. formally responded to these research abuses by passing the National Research Act in 1974, which created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. In 1979 (decades after the Nuremberg Code and the Declaration of Helsinki) the National Commission prepared a document known colloquially as the *Belmont Report*. Its constructs were formally implemented in Federal regulation in 1981 by the Depart-

ment of Health and Human Services (45 CFR Part 46) and the Food and Drug Administration (21 CFR Part 50 and 56). These regulations define requirements for research involving human subjects that apply to individual researchers and their institutions. They also established the requirements for Institutional Review Board (IRB)[1] oversight for entities conducting federally-funded research involving human subjects. In 1991, part of the regulation (45 CFR 46 Subpart A) was adopted by fifteen other U. S. federal departments and agencies in their respective regulations in what is known as the *Common Rule*.

A.2 Motivations for Ethical Guidelines in the ICTR Context

The rich field of Ethics offers mechanisms to consistently and coherently reason about specific issues, but those need to be extended and coalesced into a practical framework for ICTR. Manifold *codes of conduct* and *ethical guidelines* are available from professional organizations and societies, many of which are familiar to ICT researchers (see Appendix F.1). In aggregate they offer a solid conceptual foundation for ethical research but individually fall short of providing complete, pragmatic and applied guidance guidance in preparing research protocols or evaluating ethical issues in ICTR.

Moreover, relative to our institutionalized and socially internalized understanding of harm related to physical interactions with human subjects, our abilities to qualitatively and quantitatively assess and balance harms and benefits in ICTR are immature.

To illustrate, while there is no direct analog between biomedical and ICTR abuses it is possible to compare physical versus cyber (i.e., virtual) environments along two dimensions: the impact of harm resulting from research activities and the conditions necessary for that harm to occur. For example, in biomedical research the researcher may wish to draw blood from a subject to study the effect of an experimental drug. This requires that the subject is physically present in the research lab and that he manifests consent by signing a written form. The number of subjects is typically on the order of hundreds or thousands and the research proceeds at human speeds (i.e., the time necessary to explain the research protocol, read and sign the consent form, draw the blood, etc.). Quantitatively, the risk is often proportional to the number of subjects involved, meaning that it able to be bounded to the research subjects themselves. With computer security research, however, millions of computers and an equal or lower number of humans using those computers may be implicated as the subject of research, whether directly or indirectly. Although a researcher may likely not interact with or be able to identify those humans, if the research causes those millions of computers to crash or to reveal information about their users, the manifestation of harm to humans and damage to systems and data is foreseeable and unpredictable.

A.3 The Characteristics of ICTR and Implications for Human Subjects Protection

Traditional human subjects protections arose in response to medical research abuses in contexts that predate contemporary, ICT-pervasive environments. Research involving ICT poses challenges to stakeholder identification (i.e., attribution of sources and intermediaries of information), understanding interactions between systems and technologies, and balancing associated harms and benefits. ICTR is largely driven by needs for empirical data about the technical functioning of and

[1] While *Institutional Review Board* (IRB) is the name used in the United States for ethical review bodies, we use the more general term *Research Ethics Board* (REB) in this document.

human interactions with information systems and networks to enhance knowledge about information security and network management. Examples include countering ICT threats and vulnerabilities, protecting infrastructures, improving algorithms, creating new applications and architectures, optimizing traffic routing, and increasing national and international security.

The environment of research involving ICT differs from the "offline" environment that naturally lead REBs to assume a human-centered approach to the application and interpretation of ethical principles. The presence of ICT in research activities results in situations of greater scale and more dynamism than those research activities involving direct interactions with human beings. Datasets can be massive and can be combined easily. Automation greatly increases the speed with which actions and effects occur. Communications and actions are decentralized and distributed. ICT systems are composed of tightly coupled subsystems that are difficult to manipulate in isolation, can be can be complex, and are opaque. Additionally, researchers must contend with dynamic and malicious threat vectors. Research not involving ICT may implicate a few of these characteristics, while the involvement of ICT often raises many of them *at the same time*. These characteristics taken alone do not justify ICTR exceptionalism, however their confluence within ICTR presents new challenges in ethical evaluation and oversight.

For example, research involving "de-identified" data (biological samples) about humans in the context of large-scale, data-intensive research biological samples has raised active debate in the biomedical research community over the interpretation of Common Rule terms *Human Subject*, *living individual* and *about whom*. "[It] is not clear to what degree humans whose information is involved in such research are *research subjects* that trigger the regulatory protections. Under the definition in the Common Rule, they are not research subjects." [50] ICTR shares with the biomedical/behavioral community similar struggles regarding *non-human subjects research* [8]. For both domains, "[the] only barrier between the de-identified research data and it becoming *private information* as defined in the Common Rule is an act of re-identification." [8] The risks in ICT research extend beyond just identification of humans, however, to research interactions with information systems in ways that could result in harm to humans. Understanding this complex mix of risks complicates not only the application of the principle of Respect for Persons, but also the calculation of benefits and harms necessary to apply the principles of Beneficence and Justice.

It is unrealistic to expect rigid and unanimous agreement over the interpretation of the terms defined in the Common Rule. In fact, the problem of vaguely defined terms in the Common Rule vis-a-vis effective ethical review has been discussed before. [24] The underlying objective of protecting humans (be they direct subjects, non-subjects, or potentially harmed parties through misuse or abuse of information or information systems) remains. While the application of core ethical concepts for any given case may be fact-specific, the objective can be achieved by adhering to a framework for consistent and transparent application and evaluation of ethical practices. Common use of such a framework helps build trust that ICT researchers are adhering to traditional standards set forth in the Common Rule. This framework enables oversight entities that may not be familiar with the ICTR domain to understand and reason about the ethical issues and risk-benefit analyses as it relates to humans impacted by information and information systems. The *Menlo Report* [17] outlines such a framework and is assistive in managing ethics application challenges arising from the differences between the traditional and ICT environment, briefly described below.

Scale In ICTR, face-to-face interactions between the researcher and the subject are rare. The Belmont Report pre-dates the proliferation of ICT, which means that today's multi-site studies,

or those involving large-scale electronic data collection and analysis were not contemplated by Belmont's guidance. As such, its applications are primarily focused on human-scale biomedical and behavioral research where one-to-one researcher-subject relationships were the norm. ICT research (and data-intensive biomedical studies) may involve data regarding hundreds of thousands to millions of humans. Identifying all of those humans and obtaining informed consent may be impracticable (if not impossible).

Speed Biomedical and behavioral research in the context of Belmont proceeded at human pace where research execution was non-automated. It takes time and effort to prepare a needle and inject a drug, for the drug to take effect, and for a complete set of research subjects to participate in the study. At this pace, problems that would warrant halting an experiment can be identified well before the study is completed. Research taking place in or using an ICT environment is more efficient than direct human-to-human interactions. Actions can affect millions of devices at the edge of the network in seconds, research participation is potentially simultaneous, and the materialization of harm is often immediate. This speed differential must be taken into consideration, especially to minimize harm or to mitigate unforeseen harm in a timely and thorough manner.

Tight coupling ICT resources are interconnected via networks. The drive to innovate causes a convergence of ICTs that tightly integrate or merge formerly separate and discrete systems and islands of data. For example, contemporary smart phones may contain a vast array of sensitive data, including: lists of personal contacts linking names, addresses, phone numbers, email addresses, social network accounts, and family members; passwords for financial, utilities, insurance, or investment accounts; applications allowing remote control of home appliances and automobiles; geo-spatial location tracking of family members and friends; and private, personal photographs.

This interconnectedness exacerbates the likelihood that research activity may disclose a vast amount of personal data about not only the owner of the device, but a network of associated persons. Therefore, unintended uses or accidental destruction of the data may bear significant consequences for device owners and the persons to whom it links. There may be new tensions between scientific and commercial goals as private sector collaborators or consumers of research engage ICTR directly or indirectly to achieve socially controversial commercial goals (e.g., behavioral advertising, policing copyright, or identifying the source of anonymous communications). Further, the tight coupling of data and systems heightens the risk of collateral and cascading harms posing new challenges for identifying, responding to, and mitigating potential harms.

Decentralization ICT by definition involves interdependent relationships among various technologies. Text, audio and video communications may be located in many different places and controlled by various entities, yet converge at various places and times to convey information. This decentralization can negatively impact research activities that depend on cooperation, such as obtaining informed consent and mitigating actual harms.

A related issue is varying *levels of response capacity* [16] of each autonomous entity on the internet (be it a corporation, a service provider, or an individual's home computer network). Entities with whom a researcher interacts will have varying levels of understanding, ability, or willingness to act and decentralization can exacerbate these hurdles. For example, a computer security researcher may discover harmful activity and try to mitigate it by somehow engaging the affected parties. A report of criminal activity may not be investigated, cooperation requests to inform botnet victims may not be conveyed, or users may not unanimously understand a notice of the risk conveyed to them.

Wide distribution ICT is both decentralized and highly distributed. Some sensitive data ICT users rely upon may be decoupled from their home systems or other centralized location into a remote collection of servers. Accidental destruction of the data stored at the edge of a network, or inadvertent and simultaneous transmission of data by multiple devices, could cause significant harm when aggregated across a large user population. This spatial distribution can cause research activity to negatively impact the integrity or availability of information and information systems for quantitatively large numbers of user(s) and others relying on ICT for critical services.

Opacity Traditional biomedical or behavioral research environment has a natural focus on corporeal *human subjects* research, where interactions with living subjects are very perceptible. Contemporary behavioral and biomedical research is increasingly data-centric, involving large-scale or multi-site studies that involve vast amounts of data rather than direct human interactions. Research using ICT as a facilitator of scale rarely involves computer-mediated interactions with humans rather than direct human interactions. Often the primary concern for these studies is the collection, use, and disclosure of personally identifiable information, however interactions with ICT as the subject of research may bring up concerns of compromise to availability or integrity. As the scale, decentralization, and distribution increase, so does the opacity of knowing who, how and to what degree human beings may be impacted.

With ICTR, the direct research subjects may be exclusively humans (e.g., criminals operating botnets, users of computer interface designs), but the subject could also be the ICT itself (e.g., routers in network infrastructure), or a combination of humans and ICT (e.g., a compromised computer used by several humans). The interactions with or possible impacts on humans may be highly-mediated and indirectly observable (or not observable). Because the link between impacted persons and research is opaque, researchers and oversight entities can be uncertain about the role and necessity of human subjects protections.

Opacity challenges brought about by ICT are bi-directional. Generally, users are not privy to the inner workings of applications, devices and networks. The simplicity of user interfaces hides the technical details of storage, transmission protocols, rendering, resource management and credentialing that ensure seamless usability and enhanced functionality. Many risks are not perceptible to the user, nor is it clear what system functions are normal or what is caused by researcher activity. This complicates informed consent because of the limitations of non-technical individuals to understand risks associated with ICT. Opacity may also effect determinations of information quality and reliability.

B Relationship of the Menlo Principles to the Original Belmont Principles

What do the Belmont Report principles *Respect for Persons*, *Beneficence*, and *Justice* mean within the ICT research context? Since these ethical tenets were developed to promote the protection of human subjects in the context of biomedical and behavioral research, this section aims to evaluate, clarify and modify these principles in the ICT network and security context.

B.1 Respect for Persons

In traditional human subjects protection, respect for persons encompasses two components. First, individuals should be treated as autonomous agents. Second, persons with diminished autonomy are entitled to protection. These components are often applied via the process of informed consent. Obtaining informed consent from research subjects demonstrates that participation is voluntary and

that they receive a comprehensible description of the proposed research protocol that includes risks they face, any research benefits, and how the subject will be compensated for her efforts (e.g., time and transportation costs). Subjects are able to accept or decline participation in the research project. Research subjects are informed that they can change their mind and withdraw at any time without suffering negative consequences. In some instances it may be impossible for research subjects to withdraw (e.g., when a researcher does not know who they are, if data including their personal information has already been published, etc.).

This process is played out in traditional biomedical and behavioral research with researchers typically interacting directly with research subjects in an office, clinic, or laboratory setting. The subjects are physically present, directly communicate with the researcher and are presented with a consent form. Consent forms are written in layman's language that describes the risks of negative effects like experiencing pain or the possible consequences of having participation made public, and the protections enacted by researchers to minimize harm. Once the subject understands what participation entails and accepts the risks, she acknowledges (with a signature, or in some cases verbally) her consent to participate and she is aware that she can withdraw from participation at any time without consequence to her.

Juxtapose the above scenario to the ICTR context where obtaining informed consent is complicated by three primary challenges: identifying who should consent to research; locating and communicating with the relevant parties; and explaining intangible and/or technical issues such as research risks in layman's terms.

The first hurdle requires determining the stakeholders (assisted oftentimes by ascertaining relationships with obvious stakeholders), their level of involvement, and whether they are at-risk of harm. Identification of individuals and even groups is difficult and often impossible in ICT contexts. Computing devices are intermediaries in the communication channel between researcher and research subject, often preventing direct interaction. The research *subject* may be an individual ("natural person"), a corporation, non-profit organization, or other organizational entity ("legal person"), or the ICT itself, although any potential risk of harm will likely flow to the person(s) owning, operating or utilizing the ICT.

Once stakeholders have been identified, the next hurdle is determining the means and methods to communicate with them for the purpose of obtaining consent. In ICT contexts, the at-risk stakeholders are distanced in several ways. They may be logically distant, behind layers of service, network transit, or application platform providers. They may be distanced in space, physically located in different parts of the globe. They may also distanced temporally, in that the effects from a researcher's actions may have ramifications not only immediately, but sometimes many months after the fact. All of these factors make impracticable direct and immediate communication with them, perhaps even being dependent on the involvement of service providers or ICT system owners.

The final obstacle is conveying research risks and benefits in understandable terminology so that consent can be deemed valid. Complex and opaque technologies diminish comprehension because the vast majority of users perceive only what is conveyed through the friendly graphical user interfaces which hide the complex interactions between a multitude of underlying operating systems, applications, networks, communication protocols, and technical policies. The typical user does not know what it means to have a *subverted application programming interface that allows a rootkit to conceal underlying malware*, or what risk is posed by exposing the *end-points of connection flows in stateless communications protocols*. The *levels of response capacity* [16] concept is relevant here, whereby it may be difficult for some intermediate providers or enterprise

network contact personnel to understand what the ICT researcher is explaining and to know how to properly respond to the researcher.

Overcoming these challenges to obtaining informed consent can be difficult if not impossible in network contexts. This raises legitimate debate about whether there are suitable means to achieve informed consent in certain ICTR situations, whether the idea of individual informed consent itself is an appropriate mechanism for the principle of respect for persons, or whether it is necessary to reinterpret the idea of *diminished autonomy* and advocate for the use of proxies (e.g., *legally authorized representatives*) to exercise informed consent on users' behalf when it is impracticable to identify and obtain informed consent from every human impacted by ICTR.

B.2 Beneficence

The Belmont Report specifies two general rules under the principle of beneficence that can be summarized tersely as: "(1) do not harm and (2) maximize possible benefits and minimize possible harms." Beneficence is thus applied through risk-benefit assessment, and subsequent balancing benefits to society with burdens or harms to humans involved in research.

Biomedical and behavioral research can serve as models for understanding ethical principles in ICTR, however the tensions and complexity introduced by involvement of ICT justifies some divergence from the application of those principles as envisioned by Belmont. When it was written, the internet, social networking sites, multi-user online role playing games, video conferencing, and global search engines that index data from millions of websites did not exist. It was not possible to globally monitor communication flows of millions of individuals, nor was there a risk that disrupting some random computer connected to the internet could impact physical objects or financial processes. The risks presented by these latter examples are less well understood than research that involves injecting drugs, taking blood samples, or asking sensitive questions about sexual behavior or drug use.

Privacy risks can be vexing in ICTR because of the disjointed concept of identity in relation to digital artifacts in both law and social convention. Many digital artifacts – virtual objects that reference a component of ICT – are not identifiable information that reference specific persons or render someone's identity ascertainable. ICT artifacts are not always equivalent to identifiable information as defined in other realms (e.g., HIPAA [1]). Names, fingerprints, or biometric markers are considered more reliable identifiers than internet protocol (IP) addresses or uniform resource locators (URLs). Precisely what constitutes digital *personally identifiable information* is unsettled and evolving. Network digital artifacts, such as IP addresses, may directly identify interfaces to devices connected to the internet, however those are not necessarily the origin or destination of communications. The actual sending or receiving computer(s) may be obscured by Network Address Translation devices or firewalls, or may be an intermediate computer (i.e., a router) that serves as part of the network transporting traffic along the path between sender and receiver. Even when a device maps to a specific computer operated by an identifiable person, researchers may not readily ascertain the linked identity. Obtaining the identity of an account holder associated with the IP address of a specific device at a specific time necessitates formal legal processes involving the network service provider.

Risk determinations based solely on the likelihood of direct mapping between a network artifact and an individual underestimates the potential harm to individuals. Even when researchers have made efforts to de-identify and anonymize data, their efforts may not be sufficient to pro-

tect research subjects. Many de-identified datasets are vulnerable to re-identification by linking to commercially or publicly available data. The identifiability of artifacts associated with network traffic is context-dependent. An IP address associated with darknet traffic may identify a compromised host, while an IP address in a topology trace may identify a router. Advancements in re-identification or de-anonymization techniques, combined with large research data sets and enormous amounts of commercially available data, have changed personal identification risk assessment and challenged traditional conceptions of identifiable information. For these reasons, it may be difficult or impracticable to identify at-risk populations in a network trace, such as juvenile subjects who may warrant greater protections.

While harms in the ICT context primarily relate to the identifiability of data involved in research, they can extend farther to characterize behaviors that present other risks. Individuals have both a physical identity and several virtual identities that involve their network of relationships, online purchasing behaviors, internet browsing behaviors, etc. Increasingly, the physical and virtual identities and environments intermingle as technologies, economies, and social networks advance and converge. It is possible to remotely determine where a person is located in her house by observing fluctuations in power consumption of electrical devices between rooms, or when a home is vacant or occupied. Remote control of home appliances, door locks and ignition system in vehicles, and physical tracking using geolocation of portable electronic devices all constitute potential harm to humans beyond simply disclosure of familiar identifiers such as name, address, and Social Security Number.

B.3 Justice

In the context of human subjects research protection, justice addresses fairness in selecting subjects and stakeholders and determining how to equitably apportion who ought to receive the benefits of research and bear its burdens. The presence of ICT introduces challenges to applying this principle.

In some ways, the anonymous nature of ICT decreases the problem of selection bias in that it may be difficult, if not impossible, to chose subjects based on protected attributes such as race, gender, religious affiliation, or age. Conversely, the characteristics of ICT may obfuscate the nature and scope of research harms and more importantly, who bears the burdens.

Responsible disclosure of ICT research results – most critically with vulnerability research – demands an understanding of the time-sensitivity of knowledge about vulnerabilities in terms of mitigation or exploitation of those vulnerabilities. Publication and wide dissemination of vulnerability research should be encouraged for social protection purposes, but not without considering its benefit to malicious actors. The timing and level of disclosure should not pervert the assignment of benefits and burdens by improving the ability of criminals to inflict harm upon users of ICT and place individuals and society in a position of bearing added burdens.

B.4 Respect for Law and Public Interest

The Menlo Report makes explicit an additional principle named *Respect for Law and Public Interest*. While its meaning is addressed within the original Belmont principles of Beneficence it warrants specific attention in the ICTR context due to several factors: the myriad laws that may be germane to any given ICTR; conflicts and ambiguities among laws in different geo-political jurisdictions; the difficulty in identifying stakeholders, a necessary prerequisite to enforcing legal

obligations; and possible incongruence between law and public interest (see Appendix F.2, F.3).

C Application of the Menlo Principles

At the time of the Belmont Report, ICT was burgeoning and was not integral to or embedded in our socio-economic or research realms. Research at the time was *directly* human-centered. Data collection and observations made by researchers involved direct interaction with humans that were not intermediated by computers. It was thus relatively simpler to determine who was at risk from research activities (i.e., the human research subject) and the burdens and beneficiaries of the research were likewise easier to identify. The use of ICT in research – research that relies on previously collected data stored in databases or repositories, or research targeting ICT itself and not specifically humans – introduces new and different tensions. Specifically, there are difficulties identifying from whom to seek consent, expressing and quantifying risk, and balancing benefit and harm to individuals or organizations who may not be direct subjects of research activities.

In this section we expand on the applications of principles defined in the *Menlo Report* and provide guidance on how to apply these principles to design and evaluate research that involves ICT or is *data-centered* in nature. We use *assistive questions* to guide those who are defining or evaluating research methodologies.

C.1 Respect for Persons

Respect for persons recognizes the research subject's autonomy (i.e., the ability to voluntarily participate). Implicit in its application is identification of subjects and other stakeholders in the research, and it is usually achieved by an informed consent process that includes three elements-notice, comprehension and voluntariness. Research where ICT rather than humans is the subject may still require some form of informed consent if humans can be harmed indirectly. Stakeholder analysis illuminates persons who are direct subjects of research as well as those who may be indirectly at risk from research.

C.1.1 Identification of Stakeholders

Appropriate application of the principles of Respect for Persons, Beneficence, Justice, and Respect for Law and Public Interest requires that *Stakeholder Analysis* must first be performed. This application of Respect for Persons is similarly used as an evaluative method in many fields, such as value sensitive design (a theoretically grounded approach to designing technology that considers human values in a principled manner) [21], software engineering [23], and anthropology [41]. This activity identifies the key parties affected by the research activity by way of their interests, involvement and/or their relationship (i.e., producer or recipient) to beneficial or harmful outcomes.

- **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." This includes the researcher(s), vendor(s), those who design and implement systems, and criminals or attackers.

It may be confusing to include both beneficial (positively inclined) and malicious (negatively inclined) actors in any particular stakeholders category, even though they both actively contribute to the benefit versus harm calculus. For example, it may be easier in complex criminal botnet research to derive two distinct sets of key/primary/secondary stakeholders, those who are *positively inclined* (e.g., researchers, law enforcement, commercial service providers, the general public), and those who are *negatively inclined* (e.g., virus authors, botnet controllers, spammers, "bullet-proof hosting" providers who turn a blind eye to criminal activity within their networks).

Thorough stakeholder analysis is important to identifying: the correct entity(s) from whom to seek informed consent; the party(s) who bear the burdens or face risks of research; the party(s) who will benefit from research activity; and, the party(s) who are critical to mitigation in the event that chosen risks come to fruition. For this reason, stakeholders are mentioned throughout this section.

Assistive Questions

- Can you reasonably identify and contact persons who are potentially put at risk from research activities in order to obtain informed consent?
- Can you identify the relationships between all of the stakeholders (both positively and negatively inclined) in terms of rights, responsibilities, and duties?
- Which stakeholder populations (be they groups or individuals) may experience primary and secondary effects through disclosure of vulnerabilities or disruption of operations?
- Who owns, controls, or authorizes the use of ICT resources, or the collection, use and disclosure of data related to those resources?
- Have you identified all vulnerable groups that may be affected?

C.1.2 Informed Consent

Informed consent assures that research subjects who are put at risk through their involvement in research understand the proposed research, the risks of participating and are free to accept or decline participation. These risks may involve identifiability in research data, but can extend to other potential harms.

Consent for collection of identifiable data that will be held for a long period of time for one research purpose does not translate to re-using that data for another purpose and may require re-consent. To illustrate, in traffic classification research subjects might be willing to volunteer their traffic data for research that could improve efficiency of data delivery or for situational awareness that improves national security, but not for research on optimizing revenue for the network provider or its advertising partners. Similarly, consent for academic researchers to use data in traffic classification research is not consent for an industry trade organization to use the same data to improve their peer-to-peer file-sharing profiling techniques.

In many cases, obtaining informed consent may be *impracticable* such as with children, those with incompatible language skills, or other forms of diminished autonomy. In these situation, the interests of vulnerable populations must be addressed by alternative means. A researcher may seek to obtain a *waiver of informed consent* from an oversight authority. Doing so requires the

researcher to clearly explain the risks to stakeholders and why the researcher believes it is impracticable, not just difficult or inconvenient, to obtain consent. Researchers should be prepared to show that the research involves minimal risk and/or how they are protecting stakeholders in order to justify not first obtaining their consent.

Similarly, respecting autonomy through the application of consent is complicated by the intermediated nature of ICT or by data-intensive studies, where the human who is at the other end of the device or is referenced by the data is separated from the researcher in time or space. It may be impracticable for researchers to provide notice, communicate pertinent information in an understandable manner, and assure that participation is engaged in freely. This can present an all-or-nothing decision between (a) concluding that it is impracticable to identify and obtain consent from *every human* who may be implicated by the research and seeking a waiver of consent, or (b) not doing the research. A more appropriate decision may view these challenges as the functional equivalent of diminished autonomy when engaging ICTR and apply this principle through alternative mechanisms, such as by seeking consent and cooperation of entities with existing legal relationships with end users who can serve as proxies of informed consent.

Assistive Questions Researchers should be mindful that persons' dignity, rights, and obligations are increasingly integrated with the data and ICT systems within which they communicate, transact and in general represent themselves in a cyber context.

- If the research uses/creates data, does that data reveal name, location, relations, communications or other behavioral information that could identify an individual?
- Have individuals who are identifiable in the data consented to involvement?
- Can individuals decline to participate in the research or the uses of collected data?
- What type of consent is appropriate: general and unspecified, time-limited, source-specific use, or consent for a particular type of secondary use?
- Did the purpose for using data change or expand beyond the original scope? If so, have participants re-consented to the new use?
- Are the following justifications for not obtaining informed consent present: (a) foregoing consent is truly necessary to accomplish research goals (b) all known risks are minimal (c) there is an adequate plan for debriefing subjects, when appropriate, and (d) obtaining consent truly impacts research validity and is not just an inconvenience to the researcher?

C.2 Beneficence

Beneficence focuses on the distribution of harms, benefits, and burdens of research across stakeholder populations. Research targeting ICT itself renders the harms to humans indirect and thus harder to discern, yet still potentially wide spread and immediately manifest. Data-intensive research increases the scope and immediacy of potential harms, and alters the urgency and cost of mitigation should unauthorized or accidental disclosure occur. Applying Beneficence in data and technology research raises different tensions than existed when the Belmont Report was written. For this reason, *human-harming research* rather than *human subjects research* is a more appropriate paradigm for applying ethical principles to protect persons who may be impacted by ICTR.

C.2.1 Identification of Potential Harms

Assessing potential research harm involves considering risks related to information and information systems as a whole.

Information-centric harms stem from contravening data confidentiality, availability, and integrity requirements. It also includes infringing individual and organizational rights and interests related to privacy and reputation, and psychological, financial, and physical well-being. Some personal information is more sensitive than others. Very sensitive information includes government-issued identifiers such as Social Security, driver license, health care, and financial account numbers, and biometric records. A combination of personal information is typically more sensitive than a single piece of personal information. The combination of certain types of financial information along with name, address and date of birth suggest a higher risk due to the potential for identity theft or other fraud.

Not all harms are related to data and confidentiality, but rather, involve integrity and availability of information systems. Research risks in botnet mitigation, embedded medical devices, process control systems, etc., may not involve any personally identifiable information, yet researcher actions may still involve human-harming actions from a number of other non-data related ways in which people interact with ICT.

Assistive Questions It is perhaps easiest to identify risks that are present in the face of ICT, so much so as to warrant specifically breaking them down into sections.

General Risks

- Does harm assessment consider the number of persons who may be negatively affected by research activities? While numbers can help gauge the severity of the problem consider that some harms (e.g., disclosure of sensitive information) to even a small number of persons can be serious, depending on the circumstances.
- If research depends on collection of data, is there no sufficiently similar data being collected or available?
- What is the severity of potential harms to all persons who may be affected by research activities (e.g., collection, use or disclosure of data, publication of research results, or the interaction with ICT)? Certain people may be at a higher level of risk than others.
- Have you considered unintended consequences that are reasonably likely to result from the research?
- Does the research interfere with any stakeholder's rights to access lawful internet content and use applications of his choice?

Integrity risks

- Are risks to the integrity of not only information, but also the information systems used to store and process information, considered?
- Does the research involve data quality and integrity harms such as distortion of data that may inform government policy or public perception?

Availability risks

- Have all risks to availability of information and information systems considered, including disruption due to overloading links, corruption of communication or routing pathways, or exhaustion of resources?
- Do experimental procedures take into consideration nominal use vs. transient spikes in use that may exceed anticipated peak capacity?

Confidentiality Risks

- Does the researcher plan to disclose data as part of research publication (e.g., for purposes of scientific validation) with or without anonymization or de-identification?
- Research has exposed limitations and inaccurate assumptions about the efficacy of data anonymization techniques. If harm from re-identification is foreseeable or plausible, have the the risks of re-identification been considered? Is the sole means of protection based on anonymization of a defined set of identifiers, such as those listed in HIPAA [1]? How accessible are secondary data sources that can be combined with published data to re-identify individuals?
- What are the possible risks created by the collection, use and/or disclosure of research data? Types of disclosure include: public disclosure, compelled disclosure, malicious disclosure, government disclosure, de-anonymization/re-identification, or erroneous inferences.
- If sensitive data is collected, Respect for Persons is maintained when there is no intervention or interaction with the collected data. Have you considered either not collecting sensitive data, or imposing disclosure and use restrictions?
- Does the research involve data that indirectly identifies and/or indirectly exposes a person to harm?
- How much data is involved and does the quantity of data increase the risk of identifying individuals through correlation with other data?
- Does the sensitivity of research data depend upon the methodology for collection, use, or disclosure of that data? For example, a list of IP addresses in a network topology map may not be sensitive. However, the same information from a network telescope trace of hosts compromised by specific malware may be more sensitive. While publicly available information found when performing a DNS lookup may be be deemed less sensitive, the same IP addresses when associated with traffic in botnet research may transform the level of harm associated with its public disclosure.

Secrecy and lack of transparency

- Does research involve recording or monitoring individuals' behavior or location across time and place, resulting in harms related to surveillance? Direct harms may include: identity theft, revelation of embarrassing information, government persecution, costs associated with evading surveillance.
- Does the research chill or infringe upon individual liberties or cause users to internalize research and alter their online activities?
- Does deception or lack of transparency in ICTR activities cause physical, economic, legal, reputation, or psychological harm to individuals?
- Will the research undermine cooperation from the community whose cooperation/participation is needed/targeted, or decrease cooperation with the Government?

C.2.2 Identification of Potential Benefits

Basic research typically has long-term benefits to society through the advancement of scientific knowledge. Applied research generally has immediately visible benefits. Operational improvements include improved search algorithms, new queuing techniques, new user interface capabilities.

Consent forms often explain that research participants may not receive benefits from participating in research, although volunteering to be part of research may yield results that benefit others. This is done for two reasons. First is to ensure that benefits of research are depicted as being aimed at the larger society and to materialize in the future, not to immediately benefit research participants. The burdens of research with which benefits are balanced, on the other hand, are those borne primarily by research subjects near the time of the research activity. Second is to avoid subjects being coerced into participating in research on the belief they will receive direct benefits, such as improved disease outcomes.

Sometimes benefits accrue to multiple parties, including the researcher. Vulnerability disclosure illustrates the complexity of the benefit calculus. A researcher who discovers and discloses a software or system vulnerability may benefit through recognition and publicity of her discovery. Disclosure may also benefit system owners who can take remedial action, reducing potential harm. A vendor of a vulnerable product may take legal action against the researcher, reducing her benefit. The researcher may risk legal action because she believes that the social benefit from knowing about a vulnerability outweighs the harm of being embroiled in a legal skirmish. A fair assessment of reasonably foreseeable benefits should involve all stakeholders and account for both short and long-term benefits.

Assistive Questions

- Does the research activity clearly benefit society?
- Can benefits be identified, however great or small, for all involved stakeholders?
- Can research results be immediately integrated into operational or business processes (e.g., to improve security or situational awareness), or can the research results be acted upon meaningfully by an intended beneficiary?

C.2.3 Balancing Risks and Benefits

This principle involves weighing the burdens of research and risks of harm to stakeholders (direct or indirect), against the benefits that will accrue to the larger society as a result of the research activity. The application of this principle is perhaps the most complicated because of the characteristics of ICTR. This compels us to revisit the existing guidance on research design and ethical evaluation.

Assistive Questions

- What policies and practices associated with the research methodology assure confidentiality of information?
- Is data attributable to human subjects de-identified where reasonably possible? If distinguishing between persons influences the research goals, can pseudonyms or other forms of anonymization be utilized?

- Can a researcher justify the need for the data they wish to collect – how is it relevant, reasonable, and appropriate to fulfill the articulated research purpose?
- Does the ICTR consider only collecting and maintaining personal data that are adequate, relevant and not excessive in relation to the research purposes for which they are collected and/or further processed? Could the research be conducted without the collection, use, or disclosure of the data? If ICTR involves human subjects surveillance, are minimization techniques and processes used? E.g., limited collection, purpose specification, limited data use, limited data retention, etc.?
- Is data secured, and how is it secured against threats to privacy, data integrity, or disclosure and use risks?
- When balancing harm and benefit resulting from disclosure of vulnerability information, consider which key stakeholders (positively or negatively inclined) are likely to first act upon that information? In statistical terms, how do the cumulative distribution of exploiting vulnerable systems and mitigating those vulnerabilities compare with each other, and what is the optimal time and manner of disclosing vulnerability information to maximize benefit and minimize harm?
- Who poses harmful threats and how likely will they cause harm? How much effort in terms of resources and time would be necessary for significant harm to materialize? How easy is it to obtain external sources of data for linkage? Are there open public sources, commercial sources, or foreseeable private sources of required additional information?
- What exigent circumstances should be factored into the evaluation of and justification for certain harms?
- Is there a need for empirical research within production environments? Could the same results be achieved through experimentation, either initially or completely, within simulated or isolated ICT environments?
- What controls can be considered and applied to balance risks with benefits? Examples include: using test environments, anonymization techniques, filtering sensitive data collection, limiting personal data use, implementing disclosure control techniques. Consider proven technical and policy control frameworks you may apply to control risks are: DHS Framework for Privacy Analysis of Programs, Technologies, and Applications [13]; OMB Privacy Impact Assessment Guidance [46]; and Privacy Sensitive Sharing Framework (PS2) [39].

C.2.4 Mitigation of Realized Harms

Circumstances may arise where significant harm occurs despite attempts to prevent or minimize harms, and additional harm-mitigating steps are required. ICT researchers should have (a) a response plan for reasonably foreseeable harms, and (b) a general contingency plan for low probability and high impact risks.

Consider research involving criminal botnet monitoring where a criminal gang uses a large number of personal or corporate computers for illicit purposes. Researchers may have access to or possess evidence of crimes that could include bank account authentication credentials, credit card or automated clearing house numbers, child pornography, login credentials, proprietary information, or documents containing national security or trade secrets. These researchers may take actions that intentionally or unintentionally alert the individuals responsible for crimes to retaliate

by deleting content and destroying evidence, or otherwise impairing data or systems.

Researchers should anticipate these potential events and integrate notification mechanisms into their research protocols that maximizes benefits to society, and minimizes harms and mitigates damages to known or foreseeably affected stakeholders in a reasonable time frame. This could include notifying law enforcement or other government authorities (even though such notification may not result in immediately discernible action by those who were notified). Potentially harmed stakeholders might include intermediary sites hosting a botnet command and control server, or transit providers who could be harmed by a denial of service (DoS) situation resulting from retaliation by malicious actors controlling a botnet under study, or accidental disruption of some critical assets or service by a researcher.

ICTR does not require direct human interaction to cause harm, nor direct human notification to produce benefits. Researchers have a special obligation to inform individuals or organizations whose resources and welfare may be harmed by ICTR.

Assistive Questions

- Have mitigation policies and procedures for foreseeable harms been considered in developing the research protocols? Is there a containment or response policy and can it be followed?
- Is the researcher's organization the more appropriate entity to to mitigate harm?
- Does ICTR consider risk assessment and mitigation strategies for low-probability/high-impact events?
- What is the threshold for discontinuing research in ICTR? In risky biomedical research, a Data Safety Monitoring Plan (DSMP) is created and a Board (DSMB) reviews research activities according to the plan, limiting the adverse events and occasionally urging discontinuation of excessively harmful research.
- What is the trigger that warrants notification of a breach of sensitive research data or unauthorized disclosure of personal information?
- Does the mitigation protocol consider the cause(s) and extent of the risk exposure? Is exposed sensitive data a systemic problem or an isolated incident? Did data exposure result from external malicious behavior or an un-targeted exposure? Was the data lost or stolen? Will a targeted theft result in further harmful activity?
- Has the potential for assisting negatively inclined stakeholders been adequately weighed against the likelihood of mitigating risk for positively inclined stakeholders?
- Can harm be mitigated by notifying a subset of stakeholders or proxy stakeholders when individual notification is not possible?
- Can you reference historical cases, either exemplary or controversial, in similar ICTR? Can you use these to fashion checks and balances to prevent, mitigate and respond to foreseeable harms, such as information disclosure controls and mitigation and response plans?

C.3 Justice

The Justice principle addresses who receives the benefits and who bears the burdens of research. In the language of the Belmont Report, this means that each person, or stakeholder as described in this report, receive an equal share, according to individual need, effort, societal contribution, and merit.

A primary Justice concern is the arbitrary targeting of groups or individuals based on protected characteristics. All researchers have a duty to not exclude/include individuals or groups from participation for reasons unrelated to the research purpose. The arbitrary targeting of subjects in ways that are not germane to pursuing legitimate research questions violates this principle.

Justice concerns are among the most challenging to researchers, because researchers' interests intersect with those of their subjects. "[A] reason for the IRB system is the belief that researchers have inherent conflicts of interest and other incentives that mean they cannot always be trusted to conduct research ethically without oversight. [...] In a system that trusts researchers to behave morally, and can do nothing else, researchers must internalize such values." [24]

C.3.1 Fairness and Equity

Ideally, research should be designed and conducted equitably between and across stakeholders, distributing research benefits and burdens.

The presence of ICT in research makes the application of the Justice principle both easier and more difficult. Research directed at ICT itself may be predicated on exploiting an attribute (e.g., economically disadvantaged) of persons which is not related to the research purpose. Hence, it can facilitate arbitrary targeting by proxy. On the other hand, the opacity and attribution challenges associated with ICT can inherently facilitate unbiased selection in all research as it is often impracticable to even discern those attributes.

Assistive Questions

- Does the research target certain groups by selecting research subjects based on race, sex, religious affiliation, or other legally protected attribute? If so, how does this satisfy Justice concerns?
- Is the research equitable in its treatment of all groups involved? If not, Is there a rationale for differential treatment that is clear and justifiable?
- Does the research disproportionately benefit select groups? If so, does it accrue to the detriment of others or the group(s) that shoulder the research burdens? How might it be made more equitable?
- Is there a fair and just system for appropriately compensating stakeholders who are burdened?
- If the research involves profiling, surveilling or monitoring, have researchers protected against possible uses of results for social discrimination?

C.4 Respect for Law and Public Interest

This principle has two components: Compliance with Laws and Public Good, and Transparency and Accountability. The latter aims to ensure that the research is grounded in scientific methodology, including that it is transparent, reproducible, and subject to peer review. Navigating legal compliance waters is complex. Researchers should not be expected to know or be able to interpret the myriad of relevant legal provisions. However, researchers have an obligation to inquire about legal risks and inform their research accordingly.

C.4.1 Compliance

Applying Respect for Public Interest through compliance assures that researchers engage in legal due diligence. Although ethics may be implicitly embedded in many established laws, they can extend beyond those strictures and address obligations that relate to reputation and individual well-being, for example.

Compliance with laws and regulations may prove challenging in light of their uncertain application or interpretation. ICTR is subject to domestic and foreign laws, regulations, and organizational policies. It is impractical to enumerate every relevant law for a given ICTR project. In the United States, legal risks stem from the Federal Constitution, Federal and State statutes and regulations, contract law, tort law (e.g., invasion of privacy), organizations' policies, and even industry best practices. Research design and implementation can more effectively address legal and ethical obligations by first Identifying who has legally protected interests, rather than simply asking what a law requires a researcher to do, or forum-shopping to exploit legal uncertainty.

When conducting research and assessing risk in international settings or involving stakeholders in other countries, researchers should adhere to culturally appropriate procedures or other guidelines. Similar to legal risk assessment involving domestic laws, international risk assessment may be even less clear given the discrepancies between nation-states in the substance and application of laws, rights and customs. Adherence to international ethical standards or guidelines may help mitigate research risk when the application of foreign laws are unclear or unsettled.

Assistive Questions

- If the researcher will monitor, record, or access individuals' private communications without direct consent of the communicating parties, has the collection of communications been authorized in some other way?
- Does the research involve unauthorized access to systems, networks or data?
- Does ICTR violate federal or state criminal laws, civil laws or regulations, or other nation's laws? If the ICTR conflicts with a law or regulation, is there an exception or agreement that permits research?
- What are the international and bilateral diplomatic ramifications of research activities? Should the ICTR methodology be modified or abandoned because of legal or other concerns?
- What level of discretion exists in the interpretation and application of the relevant law(s) to the specific research activities and results?

C.4.2 Transparency and Accountability

Transparency is an application of Respect for Law and Public Interest that can encourage assessing and implementing accountability. Accountability ensures that researchers behave responsibly, and ultimately it galvanizes trust in ICTR. Transparency-based accountability helps researchers, oversight entities, and other stakeholders avoid guesswork and incorrect inferences about if, when, and how ethical principles are being addressed. Transparency can expose ethical tensions, such as the researcher's interest in promoting openness and reproducibility versus withholding research findings in the interests of protecting a vulnerable population.

Transparent research should encompass and facilitate access to the following documentation:

- The name of the researchers and affiliated organization conducting the study;

- The research sponsors / partners (who provide resources to support the study);
- The objective(s) of the research;
- The target population (the population of research subjects);
- A description of the research design and methods;
- A quantitative and qualitative description of humans directly or indirectly involved (i.e., clearly indicate the method for selecting the human-centered subjects or or ICT subjects that may have human-harming potential);
- The methods by which informed consent was obtained;
- What are the potential risks and who is at risk;
- What are the potential benefits and beneficiaries;
- How are risks and benefits balanced;
- What are the risk management policies and procedures;
- The date(s) research results were published;
- The date(s), method, and scope of data collection (if any);
- The date(s) and methods of and interactions with subjects and/or ICT systems (if any);
- The duration of the study;
- For surveys, describe the format of any questions that may be presented to human subjects (i.e., the exact wording of questions asked, including the text of any preceding instruction or explanation to the interviewer or respondents that might affect the response);
- Statistical methods used in analysis (e.g. regression analysis, support vector machine);
- Fundamental assumptions (i.e., a statement regarding assumptions and caveats / limitations of the research including how unknowns are dealt with, and the risks of drawing inferences relating to unknowns); and
- Statement of impartiality (i.e., a statement disclosing research methods that ensure objectivity at all stages including question formulation, choice of fielding method, interactions and interventions with humans or ICT systems, data collection method, data analysis method, and reporting method).

Assistive Questions

- What are the attributes of the environment being studied that justify the use of the proposed ICTR methodology to achieve the stated goal(s)? Have the goals of the research been adequately described?
- Have the research design, methods and implementation been vetted by internal and/or external authorities (e.g., REBs, sponsor agency, conference program committee, program managers)?
- Are the evaluations of the risks and benefits of the research available to the public?
- Are there means by which stakeholders can request information about the research activities from researchers?
- Consider the various models of sharing and disclosure as they relate to transparency and accountability, such as *coordinated disclosure* (e.g., contacting the organization(s) identified as at-risk through the research, or a third party that can protect the vulnerable population better than the source organization), *full disclosure* (e.g., posting complete details of a vulnerability to a public security forum), *closed discussion* (e.g., reporting to a CERT, discussion in a

closed venue with other researchers, commercial vendors, and government representatives).
- Do decisions about disclosing vulnerabilities take into consideration the advantages of coordinating with affected stakeholders?
- In deciding how to responsibly disclose vulnerability information, consider factors such as: the past track record of organization responsible for the vulnerability in dealing with researchers; the severity of harm if the researcher does/does not disclose immediately; the likelihood that damaging information (e.g., software flaw) will be exploited to cause harm before those responsible for ICT can use the information to mitigate the risk.

D Synthetic Case Study

To illustrate some of the foreseeable ethical issues that ICT researchers may encounter in practice and to illustrate the above principles, we synthesize a scenario drawn, in part, from real case studies from Section F.4. To help understand how to apply the Menlo framework to this issue-laden example, consider the following questions. Who are the stakeholders involved? What reasonably foreseeable risks exist? What are the intended beneficial outcomes from the researchers' actions? How are the risks and benefits balanced? To ensure your evaluation is unbiased, it may be helpful to ascertain how the answers to these questions might be different if you were look at the situation from the perspective of other stakeholders (e.g., the owner of an infected computer) who may be unaware of the researchers' activities.

Background: In this hypothetical scenario, researchers at university in the United States begin a comprehensive study of botnet behavior. Their goal is understanding the technical, economic, and social factors that underly botnet propagation, control, and use.

- The researchers create a botnet research testbed, connected to the internet, which enables testbed machines to become infected.
- Experiences collecting and analyzing these samples indicate that attackers test their environments to determine if they are being watched. Often these tests require the infected host to perform some malicious activity (e.g., send spam, infect another host) before downloading new versions of itself or actively participating in the botnet. The researchers construct an environment is which some of these behaviors are allowed and others are disallowed.
- Further experiences show that simply running vulnerable services is insufficient to collect representative infections. The researchers create an active component of their testbed which pulls malicious content (e.g., by visiting web sites and infecting browsers, from peer-to-peer networks).
- The researchers note that many of the links and site required the user to perform the action that resulted in the infection. In an effort to better understand why users engage in dangerous behavior, the researchers construct a experiment in which similar choices are provided. Noting that users may be reluctant to visit harmful sites in a controlled environment, the researchers propose to observe user behavior unbeknownst to the end users. To accomplish this, the researchers create an experiment in which they redirect user queries for known malicious sites to a benign server they control and monitor.
- One allowed activity on the testbed is connection to a command and control server for the botnet, where commands are issued by the attacker to the botnet. By watching this channel the researchers are aware of a subset of commands issued by the attacker to their infected

machine. The commands include instances in which bots are instructed to send malicious email, initial a distributed denial of service attack, or propagate the malicious bot software to other hosts.

- During one such investigation, the researchers discover a command and control server for the botnet they are monitoring is hosted on their university's network. They obtain, through their university, access to this command and control server. The researchers monitor this server to determine botnet membership and observe how portions of the botnet are controlled.
- The researchers determine that the botnet collects files from users' computers, depositing them in a central drop location. The malware installation script has a plain text file with the location of the server, the login user name, and password. The researchers use this information to remotely log onto the server and characterize the contents of all the files collected botnet-wide. The files include personal data and copyrighted materials.
- A newer variant of this malware no longer makes this information easily accessible. In an effort to find access to future drop zones, the researchers discover several vulnerabilities in the botnet code. These vulnerabilities allow the researchers to circumvent the protective counter-measures put in place by the attacker and continue to access the drop zone.
- Analysis of the code allows the researcher to hypothesize the ability to clean up the botnet by enumerating the infected hosts, exploiting the vulnerability, and removing the infected code. While the researchers do not take action, they publicly release a proof-of-concept prototype that demonstrates feasibility of remote cleanup.
- Further evolution in the malware strain shows the attacker moved to a peer-to-peer model for command and control. Analysis of the disk contents shows the botnet has placed copies of copyrighted music as well as personal files from other infected users on the testbed disks. Researchers analyze these files to determine the extent of damage.
- To spur further botnet research and to validate the results in their publications, the researchers release copies of both the researcher-captured malicious code, and the recovered files and data, to restricted communities of security researchers.
- Analysis of this family shows a continuing trend toward sophistication in response to mitigation by the defenders. Researchers hypothesize about future techniques attackers might employ to avoid detection. The researchers suggest and publish several, yet-unwitnessed improvements, including more resilient topologies for botnet construction, novel encryption mechanisms, and cloud-based services for creation and obfuscation of the malware.
- Researchers discover that the botnet contains code capable of manipulating proprietary process control devices (e.g., SCADA, electronic voting machines, or medical devices). Researchers publish a report that clearly makes this connection as a way of demonstrating that harm in the virtual environment can today also manifest as harm in the physical environment.
- The researchers acquire one of these computer-controlled devices and reverse engineer its operation to understand the abilities of the Botnet code. They publish the capabilities of the botnet code.
- While reverse engineering the device, researchers discover a vulnerability in the device allowing remote abuse of the device's operation. The device operators do not respond to repeated queries from the researcher. The researchers publish the details of the vulnerability.

E Conclusion

This report serves as a companion to the Department of Homeland Security's *Menlo Report* [17]. To assist readers in understanding the principles described in the *Menlo Report* this document explores the historical context of human subjects protections, the motivations behind providing new guidance to ICT research, and the challenges posed by this research. The paper provides additional details on the *Menlo Report* principles and their applications. This companion includes a significant number of assistive questions to guide users in understanding the intent of the original report. A synthetic case study is provided to help readers as they explore these principles and their application.

F Appendices

F.1 Example Ethical Codes and Standards

ICT researchers may be familiar with some ethical codes and guidelines that have been developed. This section provides a survey of selected standards.

IEEE/ACM standards The Association of Computing Machinery's (ACM) Code of Ethics and Professional Conduct [5] highlights fundamental ethical considerations, specific professional responsibilities, and leadership imperatives. Section 1 entreats members to contribute to society and human well-being, avoid harm to others, not to discriminate, to be honest, and respect privacy. Professional responsibilities include instructions that ACM members obey all laws unless there is a compelling ethical basis not to, to access ICT only when authorized, maintain competence, and accept review by the organization.

The Institute of Electrical and Electronics Engineers (IEEE) maintains the IEEE Code of Ethics [28], which contains many of the ACM imperatives in an abbreviated form. 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 their work. Of particular interest are their 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 ethical codes of conduct for computer professionals. For example, IEEE and ACM have approved a joint Software Engineering Code of Ethics [6] 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 [29], Australian Computer Society, and 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.

National Academy of Sciences Some universities and research hospitals have courses on the responsible conduct of research. One textbook that is commonly used during training is "On Being a Scientist" [4] published by The National Academy of Sciences (NAS). This book provides "an overview of professional standards in research" and uses case studies as a foundation for training new scientists.

SAGE/LOPSA/USENIX SAGE, LOPSA, and USENIX issued a joint *System Administrator's Code of Ethics* [51]. This statement blends professional and ethical standards and presents them in

a straight-forward manner using brief and simple statements divided into categories. They encourage the reinforcement of the code in the mind of system administrators by providing a *diploma* style version intended to be displayed in one's office.

Responsible Disclosure Guidelines There are various formal and informal vulnerability disclosure guidelines put forward over the years [47, 44]. The National Infrastructure Advisory Council (NIAC) is one government sponsored effort that produced a Vulnerability Disclosure Framework in 2004 [47]. This framework is intended to serve a number of stakeholders (e.g., discoverers, vendors, end user organizations, and coordinators), and sets guidelines for how these stakeholder groups should act and interact. The scope of the framework is limited to discovery, mitigation and disclosure of vulnerabilities, but it does serve to show how some harms and benefits are to be balanced within the vulnerability resolution process life cycle.

Internet Advisory Board Guidelines 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* [30], 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. 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* [31], was an informational document that stressed 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." The IAB suggested that researchers attempt to "alert relevant service providers using mechanisms such as bulletin boards, mailing lists and individual mail communications." They also suggested 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 (in fact, CERT/CC no longer provides this kind of support). 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.

F.2 Examples of Relevant U.S. Laws and Guidelines

Communications Act of 1934 (as amended by the Telecom Act of 1996), 47 U.S.C. 151 et seq., http://www.fcc.gov/Reports/1934new.pdf

Communications Act of 1996, Protection of Customer Proprietary Network Information, 47 U.S.C. §222, http://www4.law.cornell.edu/uscode/47/222.html

Electronic Communications Privacy Act – Wiretap Act, 18 U.S.C. §2510-22; http://www4.law.cornell.edu/uscode/18/usc_sec_18_00002510----000-.html

Stored Communications Act, 18 U.S.C. §2701-2712, http://www4.law.cornell.edu/uscode/18/2701.html

Pen Register & Trap/Trace, 18 U.S.C. §3121-27, http://www4.law.cornell.edu/uscode/18/3121.html

Telephone Records and Privacy Protection Act, 18 U.S.C. §1039, http://www.law.cornell.edu/uscode/18/usc_sec_18_00001039----000-.html

Family Educational Rights and Privacy Act, 20 U.S.C. §1232g(a)(4)(A), http://www.law.cornell.edu/uscode/20/1232g.html

Health Insurance Portability and Accountability Act of 1996, Pub. Law 104-191, http://aspe.hhs.gov/admnsimp/pl104191.htm

Privacy Act of 1974, 5 U.S.C. §552a(a)(5), (a)(4), http://www.law.cornell.edu/uscode/5/usc_sec_05_00000552---a000-.html

Health Information Technology for Economic and Clinical Health (HITECH) Act, Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5 Feb. 17, 2009, http://www.govtrack.us/congress/bill.xpd?bill=h111-1

See generally, Smith, Robert Ellis, Compilation of State and Federal Privacy Laws, PRIVACY JOURNAL, 2002 ed. with 2009 Supp., http://www.privacyjournal.net/work1.htm (for a more comprehensive description of relevant federal laws and their state law equivalents concerning privacy and data protection).

A Notice by the Homeland Security Department, the National Institute of Standards and Technology, and the National Telecommunications and Information Administration. Models To Advance Voluntary Corporate Notification to Consumers Regarding the Illicit Use of Computer Equipment by Botnets and Related Malware. http://www.federalregister.gov/articles/2011/09/21/2011-24180/models-to-advance-voluntary-corporate-notification-to September 2011.

F.3 Examples of Relevant Foreign and International Laws and Guidelines

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OFFICIAL JOURNAL L. 281/31, Nov. 23, 1995, http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm#directive

OECD, Guidelines on the Protection of Privacy and Transborder Flow of Personal Data (Sept. 23, 1980) http://itlaw.wikia.com/wiki/OECD,_Guidelines_on_the_Protection_of_Privacy_and_Transborder_Flows_of_Personal_Data

F.4 Example Case Studies

The following case include activities both inside and outside of the academic research settings. Regardless of whether they would be subject to REB review within an academic institution, they still comprise many of the activities that regularly do occur in academic research and serve to illuminate the ethical issues.

Table 1: Case Studies and Related References

Case	Description	References
1	Learning More About the Underground Economy: A Case-Study of Keyloggers and Dropzones	[26]
2	Your Botnet is My Botnet: Analysis of a Botnet Takeover	[56]
3	Spamalytics: an empirical analysis of spam marketing conversion	[36]
4	Why and How to Perform Fraud Experiments	[33, 32, 20]
5	Studying Spamming Botnets Using Botlab	[34]
6	P2P as botnet command and control: a deeper insight	[15]
7	DDoS attacks against South Korea and U.S. government sites	[48, 53]
8	BBC TV: Experiments with commercial botnets	[38, 49]
9	Active counter-attack measures (Lycos Europe: "Make Love not Spam" Campaign, Symbiot automated "hack-back")	[14, 22]
10	Information Warfare Monitor: GhostNet	[12]
11	Tipping Point: Kraken botnet takeover	[43]
12	Tracing Anonymous Packets to Their Approximate Source	[9]
13	LxLabs Kloxo / HyperVM	[40, 57]
14	Exploiting open functionality in SMS-capable cellular networks	[19]
15	Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses	[25]
16	Black Ops 2008 – Its The End Of The Cache As We Know It	[35]
17	Theoretic advanced attack tools (WORM vs. WORM: preliminary study of an active counter-attack mechanism, Army of Botnets, An advanced hybrid peer-to-peer botnet, and How to 0wn the Internet in your spare time)	[10, 58, 59, 54]
18	Shining Light in Dark Places: Understanding the ToR Network	[42]
19	"Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces" and "Issues and etiquette concerning use of shared measurement data"	[7, 11]
20	Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT)	[3]

Learning More About the Underground Economy: A Case-Study of Keyloggers and Dropzones [26] [$Case$ 1] In order to study impersonation attacks, typically achieved using keyloggers, Holz *et al.*identified the location of *dropzones* within malware samples. These dropzones are where captured user keystrokes are sent by the malware to later be 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 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, who brokered victim notification.

Your Botnet is My Botnet: Analysis of a Botnet Takeover [56] [$Case$ 2] Stone-Gross *et al.*at University of California, Santa Barbara, analyzed the Torpig botnet by taking control of the

botnet for a brief period starting January 25, 2009. They found two unregistered domains that the botnet would use in the future, registered them first, and put up their own C&C servers using a web services provider known to be unresponsive to abuse complaints. The attackers took control back ten days after the start of the experiment, but while in control of the botnet the researchers captured over 70 GB of data collected by the bots. This data was rigorously analyzed and found to contain credentials for 8,310 accounts at 410 different financial or commercial institutions and 297,962 regular account credentials. The researchers justified their actions using two ethical principles: "The sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized, and the sinkholed botnet should collect enough information to enable notification and remediation of affected parties." Their modified C&C server also kept bots from moving off to attacker-controlled C&C servers. No attempt was made 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.

Spamalytics: an empirical analysis of spam marketing conversion [36] [$Case$ 3] Kanich *et al.*(2008) performed a study of the conversion rate of spam campaigns by infiltrating the Storm botnet and manipulating spam messages being relayed through systems the researchers controlled by altering C&C traffic. They directed users to a fake web site that mirrored web sites advertised using Storm. 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.

Why and How to Perform Fraud Experiments [33] [$Case$ 4] In this work the authors discuss their experiences conducting fraud experiments (i.e., phishing). In particular, they focus on two studies: one exploring the impact on phishing source (i.e., someone trusted versus someone random) [32] and another exploring the impact of *cousin domains* (i.e., those which sound similar to the real domain) [20]. Their purpose was not to explore these studies in depth, but rather to highlight three important ethical issues implicated by these experiments. First is the issue of *informed consent*, centering on whether it is ethical to perform the study without knowledge and consent of participants. The question here is whether the value of the study using deception outweighs the risk of users changing their behavior if they know they are being phished. A similar set of arguments are used in discussing the second issue, *deception*. Lying to users must be done with the utmost care, be overseen by a full REB committee, and should generally be avoided by researchers. Third and finally is the issue of *debriefing* (i.e., 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.

Studying Spamming Botnets Using Botlab [34] [$Case$ 5] John *et al.*(2008) researched spam-generating botnets through analysis of email messages identified by email filters at the University of Washington (UW). Using a botnet monitoring architecture incorporating malware analysis and network behavioral analysis, they were able to develop several functional defenses. They were 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 [15] [$Case$ 6] 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. [37, 27] 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.

DDoS attacks against South Korea and U.S. government sites [$Case$ 7] On July 4, 2009, government web sites in the United States and South Korea came under DDoS attack, drawing immediate press attention and concerted efforts to mitigate the attacks. On July 12, 2009, Bach Khoa Internetwork Security (BKIS), centered at the Hanoi University of Technology (HUT), announced on their blog that they received a request for assistance from the Korean CERT (KrCERT) and information that allowed them to identify eight botnet C&C suspected of controlling the DDoS attacks [18]. BKIS claimed they "fought against C&C servers [and gained] control" of two systems located in the United Kingdom. They remotely retrieved log files and then counted and geolocated over 160,000 IP addresses around the world participating in the botnet. Public disputes erupted over BKIS' actions involving KrCERT, the Asia-Pacific CERT (APCERT), the Vietnamese CERT (VNCERT) and finally the Vietnamese government [48, 53]. A BKIS representative claimed 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 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."

BBC TV: Experiments with commercial botnets [38] [$Case$ 8] In March 2009, the British Broadcasting Company (BBC) *Click* technology program chose to perform an experiment. Unlike the situation in $Case$ 11 , direct control of the botnet was exercised. BBC staff purchased the use of a malicious botnet identified after visiting internet chat rooms. They used the 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 DDoS attack against a site willing to accept the flood; (3) They left messages on the bot-infected computers; 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 probable violation of the UK's 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 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" [49].

Lycos Europe: "Make Love not Spam" Campaign [14] [$Case$ 9] 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. **Symbiot: Active Defense** 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. [22] 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.

Information Warfare Monitor: GhostNet [12] [$Case$ 10] 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 multi-national 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." [12] 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 [43] [$Case$ 11] 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.* [43] 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* [2] 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 the property is easily destructible or concealable.

Tracing Anonymous Packets to Their Approximate Source [9] [$Case$ 12] 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.

LxLabs Kloxo / HyperVM [$Case$ 13] 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. Over 100,000 accounts were 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 [40]. Just a few days before (June 6) 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* [57]. The time line in that analysis describes attempts by the unknown security researcher going back to May 21, 2009, to explain the vulnerabilities to LxLabs staff. The researcher gave up and posted the full analysis including exploit details. Within days, multiple sites using Kloxo (including Vacert.com) were attacked by unknown parties.

Exploiting open functionality in SMS-capable cellular networks [19] [$Case$ 14] 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. 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 mitigate this problem but as the solutions appear to require a complete redesign of the cellular network they urge further investigation to protect this critical infrastructure.

Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses [25] [$Case$ 15] Implantable cardiovascular defibrillators (ICD) are implanted medical devices used to sense a rapid heartbeat and administer a shock to restore a normal heart rhythm. They 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 could identify and extract information from devices, and more importantly showed the ability to change or disable therapies (what the device does in certain conditions, e.g., the ability to deliver commands to shock the patient's heart.) The potential harm from improper disclosure could be

immediate and life threatening. As such, this research differs significantly in risks from 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 individuals. The authors intentionally explore the rationale 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 [35] [$Case$ 16] 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.

WORM vs. WORM: preliminary study of an active counter-attack mechanism [10] [$Case$ 17] 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. **Army of Botnets and An advanced hybrid peer-to-peer botnet** [58, 59] 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. **How to Own the Internet in Your Spare Time** [54] 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.

Shining Light in Dark Places: Understanding the ToR Network [42] [$Case$ 18] 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.

"Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces" and "Issues and etiquette concerning use of shared measurement data" [7, 11] [$Case$ 19] Coull et al. divulged deanonymization techniques for recovering both topology and heavy-hitter (e.g. major web servers) information from anonymized datasets. While such datasets are necessary for scientific validation of research results, researchers rely on strong anonymization techniques to protect sensitive and proprietary information about their internal networks. In this case, the authors applied their technique to three datasets, two from their own respective institutions, as well as one well-known and publicly accessible dataset prominently used in the security community. To prove the correctness of their result, they published key information about the public dataset in their paper, thus revealing internals about that researcher's network.

Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT) [3] [$Case$ 20] 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 DHS Science & Technology Directorate, the Virtual Center provides a common framework for managing datasets from various internet data providers. It formalizes a process for qualified researchers to gain access to these datasets in order to prototype, test, and improve their internet threat mitigation techniques, while protecting the privacy and confidentiality of internet users.

G Acknowledgments

This work was supported by funding from the United States Department of Homeland Security, Office of Science and Technology. It does not necessarily represent the views of the authors' and participants' respective employers or the Department of Homeland Security. The authors also wish to thank the dozen or so ICTR community members whose feedback was invaluable to assuring that this document reflects the ground truth sentiments of the professionals at the front lines of ICT research ethics. The authors would like to thank Katherine Carpenter and Sven Dietrich for generously allowing their conversations and correspondence to help in the writing of this work.

Bibliography

[1] Health information portability and accountability act (HIPAA). http://www.hipaa.org/.

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

[3] Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT). http://www.predict.org.

[4] *On Being a Scientist: A Guide to Responsible Conduct in Research: Third Edition*. National Academies of Science, 2009.

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

[6] ACM/IEEE-CS Joint Task Force on Software Engineering Ethics and Professional Practices. Software Engineering Code of Ethics and Professional Practice (Version 5.2). http://www.acm.org/about/se-code.

[7] M. Allman and V. Paxson. Issues and etiquette concerning use of shared measurement data. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 135–140, 2007.

[8] K. Brothers and E. Clayton. "Human Non-Subjects Research": Privacy and Compliance, 2010.

[9] 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.

[10] 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.

[11] S. E. Coull, C. V. Wright, F. Monrose, M. P. Collins, and M. K. Reiter. Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces. In *Proceedings of the Network and Distributed System Security Symposium*, pages 35–47, 2007.

[12] 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-Network.

[13] Department of Homeland Security. Framework for Privacy Analysis of Programs, Technologies, and Applications. http://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_03-2006_framework.pdf, March 2006.

[14] 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.

[15] 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.

[16] 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.

[17] Dittrich, D. and Kenneally, E. (eds.). The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. http://www.cyber.st.dhs.gov/wp-content/uploads/2011/12/MenloPrinciplesCORE-20110915-r560.pdf, 2011.

[18] N. M. Duc. Korea and US DDoS attacks: The attacking source located in United Kingdom. http://blog.bkis.com/?p=718, July 2009.

[19] 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.

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

[21] B. Friedman. Value-sensitive design. *interactions*, 3(6):16–23, 1996.

[22] S. Gaudin. Plan to counterattack hackers draws more fire. http://www.internetnews.com/article.php/3335811, April 2004.
[23] D. Gotterbarn, K. Miller, and S. Rogerson. Software Engineering Code of Ethics. *CACM*, 40(11):110–118, Nov. 1997.
[24] C. K. Gunsalus, E. Bruner, N. Burbules, L. D. Dash, M. M. Finkin, J. Goldberg, W. Greenough, G. Miller, and M. G. Pratt. The Illinois White Paper – Improving the System for Protecting Human Subjects: Counteracting "Mission Creep". http://ssrn.com/abstract=902995, May 2006. U Illinois Law & Economics Research Paper No. LE06-016.
[25] 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.
[26] T. Holz, M. Engelberth, and F. C. Freiling. Learning more about the underground economy: A case-study of keyloggers and dropzones. In M. Backes and P. Ning, editors, *Computer Security - ESORICS 2009, 14th European Symposium on Research in Computer Security, Saint-Malo, France, September 21-23, 2009. Proceedings*, volume 5789 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2009.
[27] 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.
[28] IEEE Board of Directors. IEEE Code of Ethics, February 2006. http://www.ieee.org/portal/pages/iportals/aboutus/ethics/code.html.
[29] Institute for the Management of Information Systems. Code of Ethics. http://www.imis.org.uk/about/codeofethics/code_ethics.pdf.
[30] Internet Activities Board. RFC 1087: Ethics and the Internet. http://www.ietf.org/rfc/rfc1087.txt, January 1989.
[31] Internet Activities Board. RFC 1262: Guidelines for Internet Measurement Activities. http://www.ietf.org/rfc/rfc1262.txt, October 1991.
[32] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.
[33] M. Jakobsson, N. Johnson, and P. Finn. Why and how to perform fraud experiments. *IEEE Security and Privacy*, 6(2):66–68, 2008.
[34] J. P. John, A. Moshchuk, S. D. Gribble, and A. Krishnamurthy. Studying Spamming Botnets Using Botlab. in Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation (NSDI '09), Apr. 2009.
[35] 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.
[36] 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.
[37] 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.
[38] S. Kelly. BBC team exposes cyber crime risk, Mar. 2009. http://news.bbc.co.uk/2/hi/programmes/click_online/7932816.stm.
[39] E. E. Kenneally and K. Claffy. Dialing Privacy and Utility: A Proposed Data-Sharing Framework to Advance Internet Research. *IEEE Security and Privacy*, 8:31–39, 2010.
[40] J. Leyden. LxLabs boss found hanged after vuln wipes websites. http://www.theregister.co.uk/2009/06/09/lxlabs_funder_death/, June 2009.
[41] S. Mascarenhas-Keyes. Ethical dilemmas in professional practice in anthropology. http://www.theasa.org/networks/apply/ethics/analysis/stakeholder.htm, July 2008.
[42] 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.

[43] R. Naraine. Kraken botnet infiltration triggers ethics debate, May 2008. http://www.eweek.com/c/a/Security/Kraken-Botnet-Infiltration-Triggers-Ethics-Debate/.
[44] National Infrastructure Advisory Council. Vulnerability Disclosure Framework. http://www.dhs.gov/xlibrary/assets/vdwgreport.pdf, January 2004.
[45] Office for Human Research Protections. International Compilation of Human Research Protections: 2010 Edition. http://www.hhs.gov/ohrp/international/HSPCompilation.pdf, 2010.
[46] Office of Management and Budget. M-03-22, OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002. http://www.whitehouse.gov/omb/memoranda_m03-22, September 2003.
[47] Organization for Internet Safety. Guidelines for Security Vulnerability Reporting and Response (v2.0). http://www.oisafety.org/reference/process.pdf, September 2004.
[48] H. Phong. Korean agency accuses BKIS of violating local and int'l law. http://english.vietnamnet.vn/reports/2009/07/859068/, July 2007.
[49] B. Prince. BBC responds to botnet controversy, Mar. 2009. http://www.eweek.com/c/a/Security/BBC-Responds-to-Botnet-Controversy/.
[50] M. A. Rothstein. Is Deidentification Sufficient to Protect Health Privacy in Research?, 2010.
[51] SAGE/LOPSA/Usenix. Unix System Administrators' Code of Ethics. http://www.sage.org/ethics/, September 2003.
[52] R. Skloot. *The Immortal Life of Henrietta Lacks*. Crown Publishers, New York, 2010.
[53] T. Son. BKIS plans to sue network security agency for defamation. http://www.thanhniennews.com/society/?catid=3&newsid=51281, July 2007.
[54] 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.
[55] D. Stevenson. Information and Communication Technologies in UK Schools: An Independant Inquiry. http://rubble.heppell.net/stevenson/ICT.pdf, March 1997.
[56] 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. In *16th ACM conference on Computer and communications security (CCS 2009)*, November 2009.
[57] Unknown. Kloxo 5.75 (24 issues) multiple remote vulnerabilities. http://www.milw0rm.com/exploits/8880, May 2009.
[58] 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.
[59] 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.