Ethics in Security Research
Which Lines Should Not Be Crossed?

Sebastian Schrittwieser
Vienna University of Technology
Vienna, Austria
Email: sebastian.schrittwieser@tuwien.ac.at

Martin Mulazzani
SBA Research
Vienna, Austria
Email: mmulazzani@sba-research.org

Edgar Weippl
SBA Research
Vienna, Austria
Email: eweippl@sba-research.org

***Abstract*—Recently, several research papers in the area of information security were published that may or may not be considered unethical. Looking at these borderline cases is relevant as today's research papers will influence how young researchers conduct their research. In this paper we discuss fundamental ethical principles and their role in recent literature. We argue that the establishment of ethical guidelines or frameworks without prior discussion and consensus in the research community probably would not lead to clarity on which lines in academic research should not be crossed.**

***Index Terms*—information security; research ethics; ethical principles**

I. Introduction

Recently, a new trend in computer security research can be observed. There are several new papers that quantitatively analyze important security issues (e.g. [1], [2], [3], [4], [5]). While many earlier works looked at threats theoretically (e.g. Thompson et al.'s famous "Trusting Trust" [6] from 1984), current researchers would probably validate their research by implementing an attack and testing it "in the wild". To some extent, this trend certainly comes from several major paradigm shifts we are facing in technology. Data moves from local storage to distributed services on the Internet, massive amount of user generated content is added to social networking sites, etc. Consolidated under the term "big data" these fundamental changes in technology usage drive the trend towards research that directly influences real people and real data.

Ethical implications in this line of research are obvious and twofold. First, we always have to think about how research results could be misused. A line from a satirical song on Wernher von Braun's attitude toward the consequences of his work in Nazi Germany on the V2 rocket says *"Once the rockets are up, who cares where they come down? / That's not my department"*. Wernher von Braun was interested in researching on rocket technology and accepted that the results of his work were used to develop a weapon. Similar to this, we have to estimate how our research could be misused. Is developing analysis methods for an anonymization network such as Tor [5] ethical in consideration of the likeliness that oppressive regimes would use the research results to deanonymize users and go after them?

Second, we have to ensure that our research activities themselves do not harm others. While the possible consequences certainly are not as fundamental to humankind such as, for example, stem cell research or other issues in natural science, we still feel the need to address these ethical questions. One important reason is that security professionals and researchers personal ethics are the discerning factor between *white* and *black hats*; we need to determine how far we can go in research. For researchers in computer security the recent success of papers such as the aforementioned are an incentive to follow along this line of research.

In this work, we want to focus on the latter type of ethical implications and aim at motivating a discussion on how research activities in the field of information security can be evaluated from an ethical point of view and how, we as a community, can establish ethical standards similar to other sciences such as medical research.

II. Recent Literature

In this section, we introduce and discuss four, in our opinion, controversial papers and their ethical considerations. We want to point our that all these papers got IRB approval and it is certainly not our intention to criticize the authors for their research. Their papers should just server as examples for controversially discussed research.

A. Spamalytics – An empirical analysis of spam marketing conversion [7]

The basic idea of this research project was to analyze the economics behind a botnet used to send millions of spam messages per day. To this end, the researches broke into a botnet, analyzed it and manipulated a small percentage of the messages in a way that the receivers actions such as clicking on links was trackable for the researchers. The authors argued that their research was ethical because they were just *"passive actors"*, *"ensuring neutral actions"* and that *"users should never be worse off due to [their] activities"*.

B. Your Botnet is My Botnet: Analysis of a Botnet Takeover [2]

This paper describes the takeover of a botnet for analysis purposes. The authors were well aware of the ethical implications of breaking into a botnet's C&C server and brought the following arguments:

- *"The sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized"*

- *"The sinkholed botnet should collect enough information to enable notification and remediation of affected parties"*

C. Pharmaleaks – understanding the business of online pharmaceutical affiliate programs [8]

In this paper the underground economics of affiliate networks for pharmaceutical products on the Internet was analyzed with the help of leaked data. At the time of research that data already was "in the wild", so the researchers used this fact to justify their analysis:

- *"[...] ethics of using data that was, in all likelihood, gathered via illegal means. [...] We justify our own choice [...] by reasoning about harm."*
- *"some [...] contents have already been widely and publicly documented. Consequently, we cannot create any new harm simply through association with these entities or repeating these findings"*

D. Is the Internet for Porn? An Insight Into the Online Adult Industry [8]

The authors of this paper analyzed the economics behind traffic trading networks for websites offering adult content and even actively participated in the business by setting up their own website with mature content. Ethical considerations were discussed as follows:

- *"Clearly, one question that arises is if it is ethically acceptable [...] to participate in adult traffic trading. [...] we believe that realistic experiments are the only way to reliably estimate success rates of attacks in the real-world"*
- *"we did not withdraw any funds but forfeited our traffic trading accounts at the end of the experiments"*

III. Fundamental Principles

At first glance, all the brought arguments for ethical justification of the introduced research projects seem to be valid and fair. We now want to discuss fundamental ethical principles and compare them to the papers and their argumentation regarding research ethics. These principles do not follow any particular ethical guidelines nor are they borrowed from other science areas such as medicine. We rather tried to derive the most fundamental principles from common sense. The reasoning is that we strongly believe that without a broad consensus across the information security community about the most fundamental basics of ethical research methods, the proposal of too detailed guidelines and frameworks would not find acceptance among researchers. In Section IV this idea is discussed in detail.

A. Do not harm humans actively

A seemingly straightforward principle is that researchers should not actively harm others. For example, writing your own malware to study user infection numbers and different dissemination strategies is obviously a bad idea. However, history has shown that in other science areas, even obviously looking principles sometimes get violated. The so-called Tuskegee syphilis experiment[1] is one of the most important cases of ethics in medical research. Started in 1932 it aimed at analyzing spread and possible treatments for syphilis. In 1947 Penicillin was found be be an effective treatment for syphilis. Nevertheless, the experiments continued for 25 years before it was shutdown on public pressure in the 70's. During the 40 years of runtime, patients were not informed about available treatments, no precautions were taken that patients did not infect others, and they were also actively given false information regarding treatment. Today, it is obvious that such a study is unethical. Doctors are not only not allowed to withhold information about effective treatment but also have to explain patients the study design. In randomized double-blind studies neither the patient nor the doctor can decide whether a patient receives a new and potentially better drug or the standard treatment. No one would withhold standard treatment as in the Tuskegee study.

Today the lines that should not be crossed in medical research are well defined (such as in the Helsinki Discords [9]) and the possible impact of unethical studies is known in detail though a large number of research scandals: medial research directly affects human lives. Arguably, the impact of research in information security cannot be compared to medial research. However, several cases throughout past years have shown that it still can have dramatical impacts on involved people. While not academic research the "Craigslist Experiment"[2] has shown the impact of unethical studies in a very drastic way and it is absolutely possible to imagine that with a similar setup privacy-impacting behavior (such as [10]) or cyber-bullying on a social network may be analyzed in an academic study.

Another problematic aspect are unpredictable effects on the analyzed systems. Often it is difficult to calculate the impact of actions performed for research purposes and harm could occur even if it was not intended. For instance, a botnet is a complex and in most cases undocumented system. How can analysis be done while assuring that the performed actions do not interfere with the system and its involuntary participants in a harmful way?

B. Do not watch bad things happening

The second principle is to not watch bad things happening without helping. In real life there is even the term "non-assistance of a person in danger". For instance, if you witness a car accident with injured people, you have the legal obligation to give first aid. At first glance, this principle seems as obvious as the first one. However, an analysis of the previously discussed papers shows how difficult it is to observe it.

The authors of the Spamalytics research [1] argued to be just *"passive actors"* and were *"ensuring neutral actions"*. It is correct that the research activities did not actively harm affected users (the first principle). Further, the authors argued that by manipulating some of the spam messages, they have done good to at least some of the receivers of spam messages.

[1] http://en.wikipedia.org/wiki/Tuskegee_syphilis_experiment

[2] http://en.wikipedia.org/wiki/Jason_Fortuny#.22Craigslist_Experiment.22

However, that is exactly the crucial point. The researcher did not prevent that still millions of real spam messages were sent over the botnet causing damage to network operators and mail service providers. The researchers knew which computers were infected, but simply watched without helping. One could argue that spam is an annoying aspect of today's email communication to which most users do not pay much attention. However, it should be kept clearly in mind that there is still a large number of people who fall for these messages – otherwise the spam business would not pay off for the sender. A 2012 report by Commtouch [11] shows that still more than 50 percent of spam messages sent worldwide advertise medicine or other pharmaceutical products, which are to a large percentage counterfeited and a major health threat. Thus, preventing spam messages from being sent probably would protect people from ordering harmful fake drugs.

In [2] the authors argued that *"damage to victims [...] would be minimized"*. The problem is that it is difficult to define *"minimizing damage"*: Ultimately, it would mean that no research is possible, because the authors of the paper would have had to take actions to shut down the botnet once they got access to it. Informing victims after finishing the experiments might not meet the principle of *"minimizing damage"*.

The next obvious question is whether not to collect certain data or discard it to avoid having all information required to inform people. Assume that we would consider the last example (botnet analysis) to be unethical, that is, we define that if we see someone is harmed by malware and probably not aware of it, we should contact him. If management decides, however, that it is still bad for business we could simply not store (or delete) the IP addresses of affected machines connects but keep all the other data. We could still do our statistical analysis for the research project but "unfortunately" we would no longer have the data required to contact the users. Would that (under the previous assumption) be considered ethical? The argument for not collecting information may be to limit the cost and security concerns because identifying data must be secured well. Deleting existing data, simply to avoid the "moral duty" of contacting people does in contrast not seem to be a good idea.

And even if it seems both feasible and responsible to inform a user that her computer is part of a botnet further challenges could occur. There might be multiple users on an infected machine and informing an arbitrary user could cause some additional harm. For instance, the infection of an office computer may have been caused by deactivating the anti-virus software, surfing to Web pages not related to work, etc. Thus informing one person could cause another person to lose his job.

C. Do not perform illegal activities to harm illegal activities

Another interesting question is wether it is unethical to harm illegal activity? – or in other words: "Is being unethical to the unethical unethical?" For example, a study wants to evaluate the effectiveness of renting botnets for spamming. Since we know from [7] that conversion rates are extremely low, it would be tempting to buy botnet resources to send spam to evaluate how well the advertised quality matches the actual performance. Even if all recipients are not real people but prepared test-email addresses as to not really harm anybody by sending them spam, an ethical problem persists: You spent research money to finance illegal activity. Would it thus be a wise choice to use stolen credit card numbers to pay the botnet rental? The credit card company will most likely revoke the payment once the card is locked thus depriving the criminals of their income. Nonetheless, the fact of using a stolen credit card by itself could be considered unethical.

In [2] the authors describe how they broke into a botnet in order to analyze it. Intercepting and modifying messages of a "legal botnet" such as distributed computing projects (for instance SETI@home [12] and Folding@home [13]) would be unethical. Is a similar activity ethical simply because it is aimed at "bad" people – though no argument of self-defense can be made? Similarly, breaking into a thieve's house "to analyze which good he had stolen" is probably a bad excuse for scholarly researcher when arrested by police.

D. Do not conduct undercover research

Law enforcement has rules defining which actions in undercover work are permitted and which not and some forms of investigation require the cooperation with law enforcement. For instance, to become a member of a group of criminals some form of joining ritual such as committing a crime to prove one's ability and loyalty may be required. In academic research, cooperation with law enforcement in not yet common in many countries. Researchers trying to understand market mechanisms of local drug trafficking cannot simply go out and sell drugs at different prices and quality to figure out price elasticity and ways of disturbing an illegal market. Besides the risk of being shot by other drug dealers, their research would be illegal. Similarly, "testing" illegal markets by buying botnets or stolen credit card numbers may at least be considered unethical since bad guys receive money.

In [14] the authors argued that they *"believe that realistic experiments are the only way to reliably estimate success rates of attacks in the real-world"*. However, this reasoning does not solve the ethical dilemma. "We had to do it in that way" is never a good argument in scientific research. Nobody forces you to perform a particular research experiment. The introduced research clearly is undercover work which could lead to – at least – problematic issues regarding ethics.

IV. Discussion

On the one hand the information security research community is well aware of ethical questions within their field. Most papers dealing with large amounts of user data or breaking into systems include an ethics section and at least in the US, universities have institutional review boards where researchers must have their proposals checked. Just recently the European Union introduced an optional review process for the European grant program FP7 [3] that is to some extent comparable to IRBs

[3] http://www.surveilleadvisoryservice.eu

in the US. On the other hand, however, the comparison has shown how difficult it is to fulfill even the most fundamental ethical principles. The question that arises is how we, the information security community, can reach a more satisfying situation. Can the proposal of some kind of ethical framework help to make research ideas easier to evaluate regarding ethical aspects? We are at least skeptical on that.

One reason is that things are changing fast in information technology – much faster than in other areas. We believe there is the threat of having guidelines that do not reflect the actual technological environment. A look at the recent history of medial research shows the dilemma. Every newly developed research method raises new ethical questions that – in some cases – entail years of discussion among the community and further (i.e. politics, religion, etc.). One of the most prominent examples from recent years is the stem cell controversy which started 15 years ago with a groundbreaking work by Thomson et al. [15]. Today, the debate is still ongoing without a broad consensus in sight. Clearly, research methodologies in information security can hardly get that controversial with influences from government policy stances and religious views. However, changing research paradigms through new technological possibilities can still lead to broad and lengthy discussions hindering the adaptation of guidelines. For instance, the debate on privacy in social networks is a passionate one and unlikely to ebb out in the near future. How should an ethical guideline rule research activities dealing with large amounts of personal data from social networks when there is no broad consensus about it in the community?

Another problem that we see is the lack of discussion. At the moment, dealing with ethical questions means in most cases getting an IRB approval and justifying the research by dedicating a section to it in the paper. Ethical considerations are often seen as a necessary evil that stands between the author and his research and not something that should be taken for granted. A more open discussion on ethical aspects of our research would be desirable. Working groups such as the one that resulted in the Menlo Report [16], [17] are definitely a step in the right direction.

V. Conclusions

Similar to other sciences, in information security research the gap between what is technically possible and what is acceptable from legal and ethical point of views is huge. With this gap it is difficult to find the right place to draw the lines that should not be crossed.

In this paper, we tried to define four fundamental ethical principles that should not be violated for obvious reasons. A comparison with recent literature, however, shows how difficult it is to obey them. While we do not believe that the introduced research was ethically unacceptable (after all, the authors got IRB approval), we strongly believe that the results of the comparison shows how difficult it is to define absolute generally accepted and universally valid principles.

We believe that these questions should be actively discussed in the future, hopefully leading to similar ethical standards as we have in medical research and other natural sciences.

Acknowledgments

The research was funded by COMET K1 and grant 826461 (FIT-IT), FFG – Austrian Research Promotion Agency.

References

[1] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamalytics: an empirical analysis of spam marketing conversion," *Commun. ACM*, vol. 52, no. 9, pp. 99–107, 2009.

[2] 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 *Proceedings of the 16th ACM conference on Computer and communications security*. ACM, 2009, pp. 635–647.

[3] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[4] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 551–560.

[5] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and D. Sicker, "Shining light in dark places: Understanding the tor network," in *Privacy Enhancing Technologies*. Springer, 2008, pp. 63–76.

[6] K. Thompson, "Reflections on trusting trust," *Communications of the ACM*, vol. 27, no. 8, pp. 761–763, 1984.

[7] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage, "Spamalytics: An empirical analysis of spam marketing conversion," in *Proceedings of the 15th ACM conference on Computer and communications security*. ACM, 2008, pp. 3–14.

[8] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. Voelker, S. Savage, and K. Levchenko, "Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs," in *Proceedings of the 21st USENIX conference on Security symposium*. USENIX Association, 2012, pp. 1–1.

[9] J. Kimmelman, C. Weijer, and E. Meslin, "Helsinki discords: Fda, ethics, and international drug trials," *The Lancet*, vol. 373, no. 9657, pp. 13–14, 2009.

[10] S. Plc. (2007) Sophos facebook id probe shows 41% of users happy to reveal all to potential identity thieves. [Online; accessed 07-February-2013], http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html.

[11] Commtouch, "Internet threats trend report," 2012.

[12] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, "Seti@home: an experiment in public-resource computing," *Communications of the ACM*, vol. 45, no. 11, pp. 56–61, 2002.

[13] A. L. Beberg, D. L. Ensign, G. Jayachandran, S. Khaliq, and V. S. Pande, "Folding@ home: Lessons from eight years of volunteer distributed computing," in *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1–8.

[14] G. Wondracek, T. Holz, C. Platzer, E. Kirda, and C. Kruegel, "Is the internet for porn? an insight into the online adult industry," in *Proceedings (online) of the 9th Workshop on Economics of Information Security, Cambridge, MA*, 2010.

[15] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones, "Embryonic stem cell lines derived from human blastocysts," *science*, vol. 282, no. 5391, pp. 1145–1147, 1998.

[16] D. Dittrich and E. Kenneally, *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*, US Department of Homeland Security, 2011.

[17] M. Bailey, D. Dittrich, E. Kenneally, and D. Maughan, "The menlo report," *Security & Privacy, IEEE*, vol. 10, no. 2, pp. 71–75, 2012.