An Ethical Analysis of White-Hat Hacking

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Owen Mitsinikos

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kent Wayland, Department of Engineering and Society

Introduction

Cybersecurity is an ever-changing field in Computer Science. As brilliant minds in the cybersecurity field begin to counter an attack, hackers will find more attacks to penetrate secure systems for immoral causes. With all of these changing attacks, one type of attack has remained pretty much the same in cybersecurity's lifetime, social engineering attacks. A social engineering attack occurs when a hacker uses things outside of computers to attack a vulnerability, for example calling someone pretending to be an IT worker asking for a password, or digging through someone's trash to find a discarded notecard containing information used to facilitate the attack.

Social engineering is very hard for companies to defend against. Unlike viruses or malware, the IT specialist cannot just patch a vulnerability to keep attackers at bay. The only way to really defend against social engineering is to spread awareness of potential threats that can occur. A common way for this to happen is for a company to hire a "white-hat" hacker. A white-hat hacker is a person in the cybersecurity field who hacks companies without the intent to commit a crime. Many businesses will hire white-hat hackers to test their security, as well as to see how well the human element of their defenses hold up. While white-hat hackers have good intentions in helping the company, they often deceive the unaware employees in a way that may be considered unethical. My research paper will look into the ethics of performing a social engineering attack when white-hat hacking, specifically in regards to how it affects the everyday worker.

Background

Social engineering attacks on technological systems have existed since computers were invented, as they don't rely on any specific technological requirement to happen, just good social awareness on the attacker's end. All social engineering attacks generally follow a specific pattern called the Social Engineering Attack Framework. The Social Engineering Attack Framework consists of six phases: Attack Formulation, Information Gathering, Preparation, Develop Relationship, Exploit Relationship, and Debrief (Mouton, p. 3). The Attack Formulation phase consists of choosing a goal and victim or group of victims that will be exploited in order to attain the goal. In the Information Gathering phase, the hacker gathers public and private information and assesses whether the information is relevant and will increase the probability of forming a relationship with the victim or not. The Preparation phase has the hacker combine all of the information and develop an attack plan for obtaining the goal. The Develop Relationship phase consists of connecting with the victim through a medium and establishing trust with them in order to succeed in the next phase, Exploit Relationship. In this phase, the hacker manipulates the victim into a desired emotional state, and then elicits the information from the victim. The final phase is the Debrief phase, in which the hacker resets the victim to their original state of mind and analyzes whether they need more information from the victim or not. All of these phases are integral to performing a successful attack.

A useful thing that the Social Engineering Attack Framework can help analyze is where in a social engineering attack the victim has control to prevent it. The first three phases; Attack Formulation, Information Gathering, and Preparation all consist of the attacker gaining information on the company and employees that is already public. Companies can't defend against this, as information sharing is integrated so much in our society through social media and things like LinkedIn that it is impossible to hide potential vulnerabilities in the research phase.

The fourth phase, Develop Relationship, is also very difficult to defend against, since we as a society form and develop relationships all the time in order to improve ourselves and the sociotechnical systems around us. While being completely guarded and not interacting with anyone will prevent social engineering attacks, it is impossible to expect this of employees who are constantly trying to improve society as a whole. The last two sections, Exploit Relationship and Debrief, are the only sections of the Social Engineering Attack Framework that can be reasonably defended against. Exploit Relationship can be defended by picking up red flags on things being asked by the attacker, while Debrief can be defended against by seeing through the attacker's manipulation and realizing that they were exploiting the relationship that they built with you. While these are the two phases where the attack is out of the hacker's hands, it is still very hard to pick up on the red flags and manipulations, which is why we need white-hat hackers to train employees to get better at seeing through the attackers.

Literature Review

There have been several controversies surrounding white-hat hackers and "white-hat hacking". The biggest controversy that has plagued it for it's entire lifetime is the fact that training people to be ethical hackers can backfire, potentially turning people into knowledgeable black-hat hackers. As Vishnuram et al. (2022) put it, "it's like giving a loaded gun to a person without knowing his background and intent".This source also shows that there is controversy between ethical theorists in whether white-hat hacking is ethically right or wrong, with Kant theory saying that it is ethically wrong, while Consequential theory disagrees.

An important distinction to make when studying this field is the difference between hackers and testers. Votpka et al. (2018) performed a study that compared the two in

vulnerability testing. While I've already defined what a hacker does, a tester is someone who is hired by the company to try and catch vulnerabilities before they push the project out of the developmental cycle. Conversely, hackers are contracted once the project is being used publicly. There is an ongoing trend to try to limit the amount of hackers needed while bolstering the number of testers a project has to try to stop these controversies while also saving money in the process. Hiring testers is more efficient for companies because it is significantly cheaper to catch a bug or vulnerability early on in the production cycle rather than when the product is already pushed, due to the product having fewer users and existing infrastructure that the company has to think about working around. Testers are also hired to test many different projects a company may have, while hackers are contracted to focus on one thing. Unfortunately, this distinction limits testers in how well they can check for vulnerabilities. Votipka et al. states, "Access to the development process is a mixed blessing. Access facilitates reporting for testers by building rapport and shared language, but "outsider by design" status allows hackers to recognize mistaken assumptions." White-hat hackers are so important in the computing world due to them being an outside set of eyes whose sole purpose is to exploit the vulnerabilities. Although increasing the importance of testing is the current goal, the paper found that the hackers were generally more confident in their abilities to find vulnerabilities in systems. The paper also had the key finding, "Hackers are exposed to a wider variety of programs and vulnerabilities through the different types of employments, exercises, and communities they are involved in and the more diverse bug reports they read. This provides hackers an important advantage over testers (Votipka, p. 12)." The paper concludes by determining, "While improving testers' vulnerability-finding skills could meaningfully improve security, companies will likely still need [hackers] to find the most complex problems (Votipka, p. 13)."

Methods

In order to further my understanding of the field and the ethics around white-hat hacking, I decided to look into the Association for Computing Machinery's (ACM) Code of Ethics and Professional Conduct (2018). These rules give clear guidelines of actions that are ethically moral in the eyes of engineers. The Code of Conduct is broken up into four parts: General Ethical Principles, Professional Responsibilities, Professional Leadership Principles, and Compliance with the Code. I found that the first section, General Ethical Principles had the most information that was pertinent to my research, although there were a few important points from Professional Responsibilities and Professional Leadership Principles. Although most of the Code of Conduct references physical code, I believe that it can be referenced in terms of social engineering, as it is a field that still falls under the computing umbrella, even though it uses social vulnerabilities. After reading through the Code of Conduct, I can analyze the ACM's stance on white-hat hacking and use it to determine if white-hat hacking is ethical.

I will also analyze University of Virginia's Institutional Review Board's (IRB) Guidelines around psychological studies that deceive and withhold information from participants with the ACM's Code of Conduct. I'm doing this due to the IRB's Guidelines already being considered ethical in the eyes of the public, which will be helpful in having a grounded reference that I know is already acceptable.

Results/Discussion

As I noted in Methods, I mainly used the Association for Computing Machinery's Code of Conduct to research my question, which all of the quotes in the following sections are from.

The section that has the most information that is useful to my research was the General Ethical Principles section. These principles “outline fundamental ethical principles that form the basis for the remainder of the Code.” The first principle that I will be discussing is one of the most important, “1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.” This principle frames the problem very well, as it shows that the people are of the utmost importance when working on a project related to computing. The Code of Conduct states, “An essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security, and privacy.” I see this principle as both a stance for and against white-hat hacking. White-hat hacking has been shown to minimize negative consequences of computing, due to being able to look at a system from an attacker’s point of view and point out flaws that would not have been found. On the other hand, white-hat hacking can be seen as a breach of personal security, as you are purposely attacking employees that do not have your consent, no matter how well-intended the attack may be. I’ll discuss more on the intention in the next section, as it is an important part of the Code of Conduct. Although Section 1.1 can be seen as denouncing white-hat hacking, I believe that it has a stronger case for validating white-hat hacking, as the white-hat hacker is not doing anything malicious with the personal information of the employees, and is instead using it to bolster the personal security in the company.

Section 1.2 of the Code of Conduct is the most simple, but the most important in understanding the ACM’s thoughts on white-hat hacking. Section 1.2 states, “Avoid Harm”. While very simple when you look through the Code of Conduct at a glance, the elaboration about what is harmful, what the intentions are, and how to minimize harm are keys to the ACM’s stance on white-hat hacking. The section states, “"Harm" means negative consequences,

especially when those consequences are significant and unjust." Some examples of harm that the section delves into that I believe are useful to my research are unjustified mental injury and unjustified damage to reputation. Although these are the two interests in the text, it also states that the list of potential harm is not exhaustive, so there may be more definitions that fit the argument better. While these other definitions exist, I am only going to use the definition of harm that is stated in the text, so that I do not have to guess if the ACM believes that it is harmful or not. The interesting part of this definition is that the ACM believes that the harm has to be justified in order to be seen as ethical. Using a utilitarian approach, the harm can be justified. While the system being broken into and the information being protected differ on a case-by-case basis, the company paying the white-hat hacker to perform the hack shows that the company believes that it is important for the security of the company. Due to any number of the victims of the attack being a subset of the company, there will always be more people benefitting if the white-hat hacker performs the hack, as the needs of the company outweigh the harm done to the victims. Section 1.2 also states, "Well-intended actions, including those that accomplish assigned duties, may lead to harm. When that harm is unintended, those responsible are obliged to undo or mitigate the harm as much as possible." This quote shows that the ACM expects the white-hat hacker to debrief with the employees and company afterwards, in order to have mutual understanding of the goal of the attack as well as to clear up any harm that might have happened to the victim's mental state or reputation. Finally, the section clarifies, "When harm is an intentional part of the system, those responsible are obligated to ensure that the harm is ethically justified." This part of the section is very pertinent to my research, as it is the basis that the research is founded on. The goal of this paper is to look into whether performing a social engineering attack for security reasons is ethically justified or not, which in turn will help or hurt

white-hat hackers trying to avoid harm depending on the findings. Overall, I found that Section 1.2 does not interfere with white-hat hacking being unethical from a utilitarian perspective, since the practice of white-hat hacking is justified when determining the security needs of the company.

Although a utilitarian approach justifies Section 1.2, there are several detractors of this section that call into question inconsistencies that it has in the Code of Conduct. For example, Kantian ethics do not find Section 1.2 ethical, as Kant's Categorical Imperative states "one should always respect the humanity in others (Jankowiak, n.d.)." A virtue ethics approach also likely rejects it, due to the act of paying the white-hat hacker instead of them virtuously performing the hack due to their good will. This approach is a little more debatable, however, as the fact that the white-hat hackers are performing the hacks for the good of the company rather than malicious means may be seen as virtuous. Many members of the Computer Science community have also taken issue with Section 1.2, believing "It condones intentional harm too broadly and does not oblige those responsible to seek external justification (Becker, 2018, p. 1)." While this is important to look at and monitor if any changes happen to the code, this paper is not determining whether the ACM's Code of Ethics is ethical or not.

The last two major section of the Code of Ethics for my research is Section 1.6, which is "Respect privacy". While these sections are not as clear cut in answering my research question as the previously mentioned sections, I have found that they are still useful in the overall goal. Section 1.6 discusses the importance of privacy when working with personal information. The section states, "Computing professionals should only use personal information for legitimate ends and without violating the rights of individuals and groups." This part of the section is

somewhat paradoxical with respect to my research, as I am finding out whether white-hat hacking can be seen as “legitimate ends”.

Other sections that I found to be loosely connected to this question were Section 1.3, “Be honest and trustworthy”; Section 1.7, “Honor confidentiality”; Section 2.7, “Foster public awareness and understanding of computing, related technologies, and their consequences”; and Section 3.1, “Ensure that the public good is the central concern during all professional computing work.” I’ve grouped these all together as they are all mainly concerned with the intent of the computing professional and what the white-hat hacker does after a hack, rather than the substance of the hack itself. For example, Sections 1.3 and 1.7 mainly discuss the computing professional’s responsibility to be a good actor in the workplace, which is assumed if they are certified as a white-hat hacker. Similarly, Sections 2.7 and 3.1 look into the value that the computing professional is giving to the public. My question is not looking into how valuable white-hat hacking is to the public, as I’ve already stated in the Literature Review having an outside source test a company’s system is already known to be valuable in the professional world.

Regarding when deception is appropriate in a study, UVA’s Institutional Review Board cites three sections from the American Psychological Association’s Code of Ethics. The first section they cite is Section 5.01, “Avoidance of False or Deceptive Statements”. This section outlines that psychologists that are running studies should not deceive the public when advertising the study. It also states that they should be honest about their credentials. This can’t really be compared to white-hat hacking, as the only comparison would be with the hiring process of white-hat hackers, which is not what I am looking into. The second section cited is Section 8.07, or “Deception in Research”. This section contains many useful comparisons to

white-hat hacking. The first part of the section states, “Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's significant prospective scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible.” This section is the crux of the comparison because it states that the psychological study, as well as white-hat hacking, must have value where other methods that are not deceptive do not work. The key to determining the legitimacy of white-hat hacking in the psychology connection is to compare it with other methods of testing vulnerable systems. This brings the argument back to the ethical hackers vs. software tester debate that I outlined in the Literature Review section of this paper. As analyzed in the Literature Review section of this paper, the lack of access that white-hat hackers have gives them a very important perspective on the company’s systems that the testers do not have due to their position in the company. I believe this, as well as white-hat hackers still existing even with testers being cheaper and more efficient in the development cycle, to be enough evidence that there is an applied value to white-hat hacking that simple testing does not provide. The final section of UVA’s determination on when deception is appropriate is Section 8.08, which is the need to debrief the participants of the study. The section states, “Psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware.” Although ethical hackers must report their findings to the people that contracted them, in this comparison they should also debrief all of the employees that unknowingly participated in the process of the ethical hack. This point also calls back to Section 1.2 of the ACM’s Code of Conduct where it said, “When that harm is unintended, those responsible are obliged to undo or mitigate the harm as much as

possible." From all of my analysis of white-hat hacking, the most important thing that I've found is that a debrief for everyone involved is absolutely necessary when performing an ethical hack on a company. As long as this is the case, I've found that when using the framework of an IRB approved deceptive psychological study, the ACM has many similar guidelines when it comes to white-hat hacking.

Conclusion

Overall, when analyzing both the Association for Computing Machinery's Code of Conduct and the University of Virginia's guidelines of performing deceptive psychological studies, I have found no evidence that white-hat hacking goes against the ACM's Code of Conduct when looking at it using a utilitarian approach. Both the ACM and UVA's IRB state that a deceptive practice such as white-hat hacking must be valuable compared to non-deceptive practices, which I have determined is that case for white-hat hacking. They both also stress the need for a debrief with the participants, which attempts to undo most of the harm done in the ethical hack.

These findings are very important to the field of cybersecurity, as they have determined the ACM's stance on white-hat hacking in clear terms, as well as approving it with another framework that is known to be acceptable. Cybersecurity is a young field that is constantly evolving, which makes understanding why practices like white-hat hacking are acceptable very important as it continues to grow. The next steps from this paper are to continue to train software testers to be more efficient rather than rely on white-hat hacker to test systems, as both codes state that a non-deceptive alternative is preferable. Other steps include looking into whether the ACM's Code of Ethics is ethical or not, as well as looking into white-hat hacking from a Kantian

or Virtue Ethics approach. White-hat hacking is truly ethical according to the governing bodies in computing and is an acceptable practice to rely on in order to test the security of sensitive data.

References

Association of Computing Machinery. (2018, June 22). *ACM Code of Ethics and Professional Conduct*. Retrieved from https://www.acm.org/code-of-ethics

Becker, C. (2018, October 18). *An analysis of Principle 1.2 in the new ACM Code Of Ethics*. Unpublished manuscript, University of Toronto

Institutional Review Board for the Social and Behavioral Sciences. *Deception and/or withholding information from a participant*. Human Research Protection Program, Office of the Vice President for Research, University of Virginia. Retrieved from https://research.virginia.edu/irb-sbs/deception-andor-withholding-information-participant

Jankowiak, T. (n.d.). *Immanuel Kant*. Retrieved from https://iep.utm.edu/kantview/

Mouton F, Malan M. M., Leenen L., Venter H. S. (2014, August 13-14). *Social engineering attack framework*. 2014 Information Security for South Africa, Johannesburg, South Africa.

Vishnuram G., Tripathi K., Kumar Tyagi A. (2022 January 25-27). *Ethical Hacking: Importance, Controversies and Scope in the Future*. 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India.

Votipka D, Stevens R, Redmiles E, Hu, J., Mazurek, M. L. (2018, May 21-23). *Hackers vs. testers: A comparison of software vulnerability discovery processes*. 2018 IEEE Symposium on Security and Privacy (SP), San Francisco, California.