I Just Want Your Anonymized Contacts! Benefits and Education in Security & Privacy Research

Ty Bross
School of Informatics and Computing
Indiana University
Bloomington, IN
tbross@indiana.edu

L. Jean Camp
School of Informatics and Computing
Indiana University
Bloomington, IN
ljcamp@indiana.edu

Abstract

Does participating in privacy research benefit the participant; if so, under what conditions? How do we measure the risk and benefit of participation of privacy and security research? In this paper we describe an experiment in which we requested anonymized information in the form of hashed contacts lists. The response to the request brought forward not only another example of the privacy paradox (people give away contacts for applications but would not sell them anonymized) but also brought forward the question of research as education and awareness. After evaluation our interactions, we developed a proposal for determining if there is a benefit to participating in privacy research. Is there a benefit in privacy awareness or increased security practices for participants in privacy and security research? We sketch a coordinated cross-university study to meet three goals: implement a practical collaborative partnership; investigate the value of security research for participants in terms of education; and enable evaluation of distinct benefit assessments. [1]

***Index Terms*—Ethics; Security; Privacy**

I. Introduction

Is there a benefit in privacy awareness or increased security practices for participants in privacy and security research? If so, can we measure it? Under what conditions do privacy and security research enhance the privacy or security skills of individuals? Is it possible to make research itself human-centered; given that much security and privacy research addresses our vast ignorance in the domains of security and privacy risk?

The experience underlying this white paper began when the first author created an Android trojan horse application that steals a user's contact data from the Gmail accounts associated with the device. The goal of the human subjects interaction was to obtain information about social network size and overlap in order to create a moderately grounded diffusion model, obviously without malware. We offered participants five dollars to share anonymized contact information. (The malware and the analysis of diffusion are a component of another submission.) After asking one hundred sixteen people, primarily in computing studies, we had twenty-four participants. This outcome was not expected, given the willingness to share information online. In retrospect, our expectation of rational actions on the part of potential subject participants was itself irrational. Certainly the privacy paradox has long been recognized, e.g. [**?**]. It has been illustrated experimentally and found to be somewhat consistent [**?**], [**?**], [**?**], [**?**].

In this paper we use this grounding of our interactions with individuals in order to make two arguments. First, and critically, the act of being recruited and also participating in research in the arena of security and privacy may enhance security and privacy awareness of the potential recruits. That is to say, there may be an educational benefit for study participants that would be worthwhile exploration absent the value of the research. Were we to understand this benefit, then security and privacy research could be designed to increase the benefit of the participants. Second, individuals do care about privacy. Certainly, the same people who refused to allow us to the view the contacts have a high probability of sharing exactly that information to Android applications. Indeed, the individuals who refuse almost certainly share contacts on Facebook, yet when offered $5 to share they refused to provide these anonymized.

In the next section, we describe our experiment and reference some language from the IRB to illustrate the grounding in risks and permissions discourse. In Section **??**, we argue for investigation into the value of security and privacy for research subjects, motivating this from the far more problematic medical domain. Next (Sec. **??**) we propose a possible experiment that might be implemented simultaneously at multiple universities to determine if there is a secondary benefit in participating in security and privacy research. We conclude in the following section with an explicit call for dialogue and collaborative partnerships.

II. Risks and Permissions

What risks do individuals take when interacting with their phones? In this section we provide a high abbreviated description of the risks taken by potential study participants when downloading Android apps. What information is normally provided when faced with the risk documented and emulated

[1] Ty Bross & L. Jean Camp, "Benefits and Education in Security & Privacy Research", CREDS: Cyber-security Research Ethics Dialog Strategy: A IEEE Symposium on Security & Privacy Workshop, San Francisco, CA 23 May 2013.

in our experimental framing? Note that the requests we made were in-person but far less than the data requested in Android's explicit permission scheme.

The global surge in popularity of smart phones, combined with the increased processing power, increased bandwidth, and the sensitivity of data stored on the phones, makes them a desirable target for malware developers. Cellular phones that utilize the Android, iOS, or Windows Phone operating systems, have been experiencing a surge in use over the past few years. Eric Schmidt, CEO of Google, stated in late 2012 that there were 1.3 million Android activations per day and a total of almost 500 million Android devices currently already use at that time [**?**]. We sought to explore the implications of social spread of malware. We did so by creating a malicious application with obvious monetization. We then sought to model the spread, assuming if individuals receive a text from a trusted individual recommending an app, the probability of installation is larger than zero. *Security and privacy research is taking place in an arena where potential research participants have little understanding of their privacy and security risks.*

We illustrated in the lab that embedding such code was possible shortly before the first (similar) iOS malware was reported [**?**]. The similarities of our theoretical malware and the malware in practice were remarkable. Working with a member of the complex systems lab, we sought to examine if this would be a problematic method of malware propagation. Our research agenda was as follows: 1) implement exemplar malware; 2) obtain data on social network sizes in cell phones; 3) use this data to model diffusion using different rates of transmission and network size; and 4) implement user studies of willingness to install an application from an unofficial store based on a text from a friend.

As shown in the image in the appendix, when a person attempts to install an app on her Android device, a list of the permissions required will be displayed. If the person accepts the required permissions the app is then installed and the permissions are granted to it. In this permissions model, a person must accept all of the required permissions or the app will not be installed. Chin et al. found that people depend more on social factors when installing apps than the permissions required. It is postulated that this is because the permission list that appears at install-time is explicit but difficult to read [**?**]Confidence and that many people simply do not want to be bothered to read popup notifications if they interrupt their desired task, such as app installation [**?**]. This pattern has also been observed in the behavior of people when displayed by EULAs and privacy policies [**?**].

When a person attempts to install an app downloaded from the Google Play store, a list of permissions are shown and that the person may tap or click on each required permission to see for what functionality the permission is required. When a person downloads an app from a 3rd-party source, such as a download from a website, the person is unable to expand the permission list to see what each actions each permission allows the app to perform.

Recall the first of phases of our experiment was the design and testing of the trojan horse app. The second phase was intended to consist of a survey of 100 students from a major university located in the United States. This survey collected simple demographic data and the collection of the user's contact list from their smart phones.

The data collected about the user's contacts during our experiment was limited to e-mail addresses. Other data that can be collected include name, phone number(s), physical address, photos, and what group the contact belongs to in the user's Gmail account. The primary information we were concerned with is the contact's e-mail address. We hashed each e-mail address to maintain privacy while allowing us to compare addresses for uniqueness, then deleted the unencrypted information. We had the (very simple) code available for individuals to view before connecting if they so desired, with the deletion being obvious. The results of this solicitation were remarkably negative. Individuals refused to share contact information in person despite the willingness to share them over the network to those provide apps.

In previous work, individuals have been decried for willingness to provide passwords for candy bars. The obvious criticism of these experiments is that the passwords were never confirmed. Indeed, the same experiment could indicate that individuals are willing to lie about their passwords for chocolate bars, which presents a different perspective altogether. Yet other research has shown that direct willingness to pay illustrates that many people find even twenty-five cents too much to spend [**?**] In contrast privacy is also a luxury good [**?**], where people are willing to pay for privacy [**?**].

Privacy may also be a function of awareness. That is, people may not care about privacy; people may not know about privacy risks; or people may know and care yet be unable to protect themselves. The three very different answers, and the populations and situations in which they apply, have a profound implication for the importance of privacy-centered research. In this case, the solicitation of participants may have had somebenefit. It is the perception of the researchers that all the solicited potential participants came from the discussion with an increased awareness of privacy. However, while we have no basis other than unreliable (and admittedly rather self-interested perception) the question itself came to the fore.

III. Implications

In the case of health research, individuals are forced to choose between a placebo or standard treatment, versus the experimental treatment. Given that privacy research takes place in an age of massive data compilations, what choice do our subjects face? Is it possible to quantify the value (as negative risk or positive benefit) in participating in security and privacy research? With respect to research in health, at the close of an experiment individuals are either in better health or worse health. Physical health can be evaluated to determine if the care provided was a benefit compared to the standard medical response. How might such a comparison be made in the case of privacy research?

Henrietta Lacks [**?**] most famously bequeathed to science a line of cancer cells that have been the foundation on which fortunes, fame, and of course, curing families has been built. John Hopkins provided free or reduced price medical care to Mrs. Lacks, as it did to all local African-Americans in those days of segregated Maryland. In exchange for cancer cells, of which only Mrs. Lack's proved resilient enough for laboratory growth, patients received medical care. The quality of the care was noted by the scientist who extracted Mrs. Lack's cells, who remembered that her nails had been manicured before death. Of course, denial of medical care to the unwilling participants of the Tuskegee Experiments cannot be forgotten.

Similarly, the provision of care for testing in developing countries is problematic. Due to factors beyond the control of researchers, individuals can choose between no medical care, or medical care in the context of research. Medical researchers have the Declaration of Helsinki, which defines the *Ethical Principles for Medical Research Involving Human Subjects*, for guidance. There is a core which is shared in virtual and medical domains: "In [medical] research involving human subjects, the well-being of the individual research subject must take precedence over all other interests". A component of this is understanding the potential benefits for participants. Yet there are fundamental differences. First, of course, medical care may be life or death. Our stakes are simply lower. Second, is that our subjects are often truly anonymous. For example, participants on the Tor network should not be subject to potential identification and session linking in order to provide education. Third, while much is to be learned in medical education, public education and risk communication for security and privacy is infant in comparison.

Resolution of the ethics in medical research includes educational components and auditing of the research [**?**]. "Enhancements in knowledge" as well as health benefits are a widely-recognized component of research that is ethical [**?**]. The value of health education is confirmed in the Declaration of Helsinki. Also highlighted in global medical ethics is the importance of collaborative partnerships [**?**] to ensure that abstract agreement about ethical behavior does not fragment into splintered disagreement in practice. The USACM and IEEE–USA have proposed collaborative partnership as a mode of research guidance and audit in response to the Menlo Report[2]that offers a step in that direction.

Another core issue in developing county population is benefits, with and without the research. This is related to conditions on the ground. Should privacy and security research be subject to greater or less constraints due to the state of the network? Do well-documented geographical differences in networks [**?**] correspond to different needs for research? What about temporal changes, if security violations increase should standards for researchers remain static? This is not to compare the profound human tragedy of HIV in developing countries with fatality-free but much-hyped 'cyberwar', but rather an effort to build upon the understanding from the profound ethical struggles arising among the scientists involved in the deadly war against disease. Indeed, because of the lessor benefits and lower risk, many of the challenges of the networking community are far more simple, despite the existence of a physical point of contact with medical research.

No doubt networking can learn from medical ethics. In our case, not only do community standards lack uniformity, but the law of the land may prohibit actions which would appear to be optimal under the medical model (if one models patching and recovery as 'healing'). In computer security, a choice to implement automated recovery for an involuntary research subject may even be prohibited by law. If the researchers tidying the botnet in [**?**] had been recovered and patched, these researchers may in fact have committed a felony under the Computer Fraud and Abuse Act.

IV. PROPOSED COLLABORATIVE INVESTIGATION

We have identified education of participants, review of research, and clear quantification of risks and benefits as components of ethical research. In this section we propose also adopting the process of collaborative efforts to meet the goals of education, consistent review, and benefit quantification. Specifically, in this section, we suggest a series of collaborative experiments to determine if participating in security and privacy experiments has a positive effect on subject awareness of privacy. Here we outline something more than a straw proposal. Such a series of experiments could inform the development of policy with respect to security and privacy experiments online by creating an understanding of the potential benefits for the subjects. Simultaneously, this experimental approach will also illustrate possible losses should this type of research be prohibited. That the community would be willing to test the value of our own research on subjects is a bold proposal.

The experiment proposed here is intended to be a concrete step to address core variables from both personal research experience and other ethical domains: 1. review differences; 2. education; and 3. benefit assessment. The purpose of selecting and proposing explicit experiments is to move forward the debate with respect to human-subjects research beyond the theoretical to a coordinated collaborative partnership. While it is not expected that this necessarily will emerge from the discussion in the proposed form, the goal is to ground the debate in specified, actionable, next steps.

We propose a series of three identical experiments across University introductory level classes and simultaneously with subject populations in industry. In one section, the participants are subjects of experiments on privacy; each one of which is cleared with the appropriate IRB. In the second, the participants do not engage as research subjects except with temporally discrete surveys. That is, in both sections there is a survey at the beginning and the end of the class to determine

[2]DHS-2011-0074 (2011) The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research" for the Department of Homeland Security (DHS), Science and Technology, Cyber Security Division (CSD), Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT), Fed Reg v76, (249), 12/28/11

participant awareness of security and privacy risks; and ability to mitigate those threats if so desired.

Study 1: Participants are asked to price various types of information, building on the classic experiment for willing to pay or willingness to sell by Acquisti and Grosslags [**?**].

Study 2: Participants appear to evaluate two web sites, experience failures of benevolence and competence, with their responses evaluated, as first done in [**?**].

Study 3: Participants engage in experimental analyses of efficacy of password interactions, grounded in [**?**].

This set of studies combines core components of security research. In the first there is explicit pricing. The pricing in the first can inform benefit evaluation. In the second, there is apparent security and privacy risk, again providing data for benefit evaluation. The third requires self-reflection and determination of a willingness to invest in password effort. The experiments were specifically selected to include one with acceptance of visible (apparent) security harm, rather than, for example phishing, where the results would be a susceptibility not of acceptance. The three experiments were chosen to enable results suitable for a wide range of modeling of potential benefits. The three experiments were also chosen as they are from three laboratories which have made contributions to human risk behaviors with respect to security and privacy. Finally, these were selected as each is strongly theoretically grounded, and each in a distinct domain (economics, behavioral theory, and usability respectively).

In development and preparation or the studies, the researchers at each university can develop and document the dialogue with our respective IRBs (or corporate counsel as appropriate). This alone would be a contribution to consistent applications of ethical research, as the feedback we provide to our IRBs will inform our individual future research.

The purpose of the studies is to generate findings that can assist in valuing the benefits of privacy and security research, using both revealed and expressed preferences. The findings themselves, as these are repetitions of studies that have already been completed, are not expected to be particularly novel. The purpose is instead the macro-examination of the value of research participation. Recall each participant section would have a questionnaire at the beginning and the end of the semester with respect to three topics: privacy awareness, ethical awareness, and security awareness.

The goal of this proposal, which will not doubt be improved by workshop discourse, is to address the questions from the introduction. First, does participating in privacy and security research benefit the participant. Second, if so, under what conditions? Third, how do we measure the risk and benefit of participation of privacy and security research?

V. Conclusion

The effect or benefit for the individual participating in privacy and security research must be addressed for informed consensus on research practice. Are we changing our subjects when we engage in research? If so, how?

We propose to learn from the far greater ethical challenges in the medical world by taking the first step of evaluating the benefits of participating in privacy and security research. The experimental structure is proposed as first steps to explore issues of education, benefits, and consistency of practice using collaborative partnerships. More difficult questions remain; for example, this would not address issues of appropriate response to botnet participants by industry or academy, or ethical research on anonymity. By addressing benefits, educational value, and consistency of review in a series of theoretically grounded experiments (one of which includes deception) the debate as whole may be more informed. If not this, how? If not now, when?

Acknowledgment

This material is based upon work supported, in part, by the NSF under Grant CNS 1250367. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or IU.

References

[1] First ios malware hits app store. http://www.forbes.com/sites/adriankingsleyhughes/2012/07/06/first-ios-malware-hits-app-store/, 2012.

[2] N. Awad and M. Krishnan. The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, 30(1):13–28, 2006.

[3] S. R. Benatar. Reflections and recommendations on research ethics in developing countries. *Social science & medicine*, 54(7):1131–1141, 2002.

[4] R. Böhme and S. Koble. On the viability of privacy-enhancing technologies in a self-regulated business-to-consumer market: Will privacy remain a luxury good. In *Workshop on the Economics of Information Security (WEIS), Carnegie Mellon University, Pittsburgh, PA*, 2007.

[5] L. J. Camp, C. McGrath, and A. Genkina. Security and morality: A tale of user deceit. *Models of Trust for the Web (MTW06)*, 2006.

[6] E. Chin, A. Porter Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, 2012.

[7] S. Egelman, A. P. Felt, and D. Wagner. Choice architecture and smartphone privacy: Theres a price for that. In *Workshop on the Economics of Information Security (WEIS)*, 2012.

[8] E. J. Emanuel, D. Wendler, and C. Grady. What makes clinical research ethical? *JAMA: the journal of the American Medical Association*, 283(20):2701–2711, 2000.

[9] E. J. Emanuel, D. Wendler, J. Killen, and C. Grady. What makes clinical research in developing countries ethical? the benchmarks of ethical research. *Journal of Infectious Diseases*, 189(5):930–937, 2004.

[10] V. Garg, C. Kanich, and L. J. Camp. Macroeconomic analysis of ecrime in crowd-sourced labor markets: Mechanical turk vs. freelancer. In *11th Annual Workshop on the Economics of Information Security*. WEIS, 2012.

[11] J. Grossklags and A. Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *Workshop on the Economics of Information Security (WEIS)*, 2007.

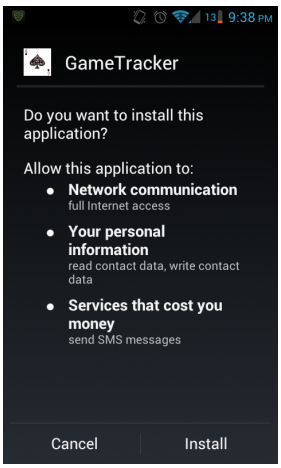
[12] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: an empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM conference on Computer and communications security*, CCS '08, pages 3–14, New York, NY, USA, 2008. ACM.

[13] H. Nissenbaum and E. Felton. Computer security: Competing concepts. In *30th Research Conference on Communication, Information and Internet Policy*, 2002.

[14] P. A. Norberg, D. R. Horne, and D. A. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

[15] A. Porter Felt, S. Egelman, M. finifter, D. Akhawe, and D. Wagner. How to ask for permission. In *USENIX workshop on hot topics in security*, 2012.
[16] E. Schmidt. 'there are now 1.3 million android device activations per day'. http://techcrunch.com/2012/09/05/eric-schmidt-there-are-now-1-3-million-android-device-activations-per-day/, 2012.
[17] A. Shostack and P. Sylverson. What price privacy? In L. J. Camp and S. Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, pages 129–142. Springer, New York, NY, 2004.
[18] R. Skloot. *The immortal life of Henrietta Lacks*. Crown, 2010.
[19] J. Tsai, L. Cranor, A. Acquisti, and C. Fong. What's it to you? a survey of online privacy concerns and risks. *A Survey of Online Privacy Concerns and Risks (October 2006) .NET Institute Working Paper*, (06-29), 2006.
[20] D. Weirich and M. A. Sasse. Persuasive password security. *Proc. of CHI n factors in computing systems*, pages 139–140, 2001.

VI. Appendix 1: Privacy Communication



VII. Appendix 2: Solicitation

Our solicitation was as follows, in written form. The actual solicitation consisted of a person with a sign offering the $5 amount for a moment of time. Since each solicitation required verbal physical interaction, this captures only the rough nature of the experiment.

The purpose of including this illustration is to show how no benefits are included. There is no educational material, or grounding with respect to the purpose. With an understanding of possible benefits, anyone who considered participation could have had some education.

> You are invited to participate in a research study of smartphone security. You were selected as a possible subject because you possess a smartphone. We ask that you read this form and ask any questions you may have before agreeing to be in the study.
> The study is being conducted by Ty Bross, with the School of Informatics and Computing, Security Informatics.
>
> STUDY PURPOSE:
> The purpose of this study is to determine the potential for automated diffusion of information through smartphone contact lists.
>
> PROCEDURES FOR THE STUDY:
> If you agree to be in the study, you will do the following things:
> Save your phones contact list and transfer it to a computer for analysis where it will be immediately encrypted. This will be performed by exporting contacts to a .vcf file from the phone and sending the file to the investigators e-mail address.
> You will also complete a brief demographic survey asking your age (but not date of birth), gender, major.
>
> CONFIDENTIALITY:

Your contacts information will be confidential. Contact information collected will be limited to encrypted versions of your contacts e-mail addresses. The method we will use is called hashing. This method will ensure that even the investigators will not have the ability to view this information. When your contacts file is transferred to the investigators computer, a script operation will be performed that will encrypt the contents of the file and output the encrypted versions to a text file. Once this is completed all copies of the contacts file in the investigators possession will be deleted. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Your identity will be held in confidence in reports in which the study may be published.
Organizations that may inspect and/or copy your research records for quality assurance and data analysis include groups such as the study investigator and his/her research associates, the Indiana University Institutional Review Board or its designees, etc., who may need to access your research records.

PAYMENT:
You will be compensated $5 US upon successful transfer of your contact list to the investigator.