Ethical Hazards and Safeguards in Penetration Testing

Shamal Faily
Bournemouth University
Poole, UK
sfaily@bournemouth.ac.uk

Claudia Iacob
University of Portsmouth
Portsmouth, UK
claudia.iacob@port.ac.uk

Sarah Field
MWR InfoSecurity
Basingstoke, UK
sarah.field@mwrinfosecurity.com

Penetration testing entails attacking a system to identify and report insecurity, but doing so without harming the system nor encroaching on the dignity of those affected by it. To improve the interaction between penetration testers and their processes and technology, we need to understand the factors that affect decisions they make with ethical import. This paper presents four ethical hazards faced by penetration testers, and three safeguards that address them. We also present preliminary results validating the hazards and safeguards.

Penetration testing; ethics; Grounded Theory; personas; goal model; CAIRIS; GRL

1. INTRODUCTION

Penetration testers attack systems to gain assurance about their security. These attacks take the form of authorised "penetration tests" that probe a system's defences; these defences are breached to evaluate the impact of any weaknesses; the results of these tests are used to improve a system's security, making them resilient to further attacks. Penetration testing requires technical prowess, creativity, and ingenuity to find unexpected ways of breaching a system. However, penetration testers face the added constraint that finding and exploiting vulnerabilities should neither harm the system nor encroach on the dignity of those affected by it.

When faced with ethical dilemmas during the planning, execution or reporting of a test, penetration testers are expected to adopt different ethical perspectives when deciding the right course of action (Mouton et al. 2015). Recent work by the authors (Faily et al. 2015) has shown that penetration testers are subject to certain decision making biases; when faced with ethical dilemmas, these biases influence decisions of ethical import. To improve the interaction between testers and their tools and techniques, we need to typify the situations where such decisions might be made, and identify factors that positively or negatively impact these situations.

In this paper, we present four ethical hazards associated with penetration testing, and three safeguards that minimise their impact. We describe the approach taken in Section 2, summarise the ethical hazards and safeguards in Section 3, and present preliminary results validating the hazards and safeguards in Section 4.

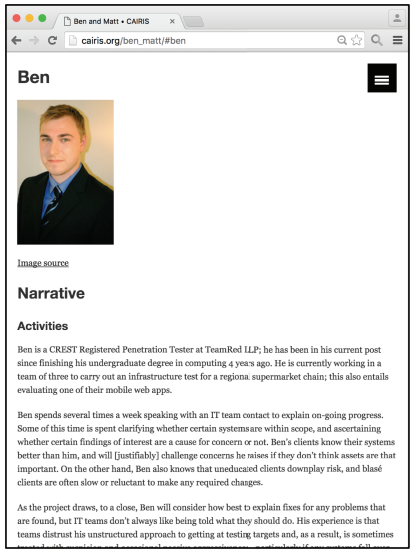

Figure 1: *Penetration tester persona (Ben) derived from the model of ethical hazards and safeguards*

2. APPROACH

We analysed transcripts from eight semi-structured interviews with professional penetration testers; each interview lasted approximately 45 minutes. These transcripts were collected during previous

work by the authors (Faily et al. 2015). Using Grounded Theory (Corbin and Strauss 2008), we analysed the transcripts and develop a qualitative model of ethical hazards and safeguards.

To validate this emerging model, we used the Persona Case process (Faily and Fléchais 2011) and the CAIRIS software tool (Faily 2016) to derive two personas from this grounded theory model. These personas were based on an experienced penetration tester (Ben) and a penetration test manager (Matt). The personas were distributed to the interviewees for comments and, based on the feedback, subsequent axial and selective coding identified several key concepts. The personas have also been made publicly available (Faily and Iacob 2015), e.g. Figure.1.

3. ETHICAL HAZARDS AND SAFEGUARDS

We elicited the following four *ethical hazards*: these are situations likely to increase the probability of unethical behaviour because of the means, motive, and opportunity to engage in such behaviour (Pendse 2011).

- Legal Ambiguity: The uncertainty associated with addressing unusual forms of illegality when encountered, or dilemmas between following the agreed rules of engagement, or informing law enforcement agencies
- Human Targets: Any testing activities with the potential to jeopardise the career or well-being of test subjects.
- Red Team vs Blue Team: Tensions that arise between testers (red teams) and client IT teams responsible for interacting with them (blue teams).
- Client Indifference: Occurrences where clients are reluctant to make changes prescribed by penetration testers, or downplay the significance of problems found.

These hazards are mitigated by the following three safeguards:

- Risk Articulation: the explanation of security risks, and the impact these have when put in a meaningful context.
- Service Comprehension: the understanding that clients have about the penetration test service they have commissioned.
- Responsibility to Practice: the sense of responsibility that testers have to the penetration testing profession.

4. PRELIMINARY RESULTS

To confirm the presence of the ethical hazards and safeguards, we surveyed professional penetration testers to compare whether their understanding of ethical hazards and safeguards corresponded with those in the model we created. Because of the sensitivity of this topic, we used the personas as a vehicle for getting testers to describe their personal opinions about these ethical issues. We applied the approach described in (Faily and Fléchais 2014) to generate a goal model of one of the personas: Ben. This model was represented in the Goal-oriented Requirements Language (Liu and Yu 2004). All four ethical hazards were evident from this model. The goal model was imported into jUCMNav (Mussbacher et al. 2009), where the model was evaluated to examine the effects that denying the goals associated with the ethical hazards might have on other goals that Ben wishes to satisfy.

We created a premortem scenario (Klein 2007) to illustrate the impact of these goals being denied. This was circulated to penetration testers in a security consultancy with approximately the same level of professional experience as Ben. The scenario described how Ben, while supervising a junior tester, carried out a test with catastrophic results to the client. Participants were asked to provide open-ended responses with reasons why Ben might have behaved unethically (ethical hazards), together with things that the company could do to address the problems found (safeguards).

Five participants responded via email with 18 candidate ethical hazards, and 21 candidate safeguards. Each candidate ethical hazard was coded based on related goals, and goals directly harmed as a result. Each candidate safeguard was categorised based on related goals, and goals directly safeguarded.

14 of the 18 candidate ethical hazards corresponded with at least one ethical hazard from the model. Three of the candidate ethical hazards not in the model related to unwarranted trust placed in the tools used by the junior tester; the other was attributed to Ben not properly supervising the junior tester. All candidate safeguards corresponded to at least one safeguard from the model.

In future work, we will replicate this validation for Matt, and a premortem scenario specific to a penetration test manager. We will also explore the value tensions associated with penetration testing tools, and the factors that influence the trust professional testers place in them.

REFERENCES

Corbin, J. M. and A. L. Strauss (2008). *Basics of qualitative research : techniques and procedures for developing grounded theory* (3rd ed.). Sage Publications, Inc.

Faily, S. (2016, May). CAIRIS web site. http://cairis.org.

Faily, S. and I. Fléchais (2011). Persona cases: a technique for grounding personas. In *Proceedings of the 29th international conference on Human factors in computing systems*, CHI '11, pp. 2267–2270. ACM.

Faily, S. and I. Fléchais (2014). Eliciting and Visualising Trust Expectations using Persona Trust Characteristics and Goal Models. In *Proceedings of the 6th International Workshop on Social Software Engineering*, SSE 2014, pp. 17–24. ACM.

Faily, S. and C. Iacob (2015). Ben and Matt: Penetration Tester Personas. http://cairis.org/ben_matt.

Faily, S., J. McAlaney, and C. Iacob (2015). Ethical Dilemmas and Dimensions in Penetration Testing. In *Proceedings of the 9th International Symposium on Human Aspects of Information Security & Assurance*, pp. 233–242. University of Plymouth.

Klein, G. (2007). Performing a project premortem. *Harvard Business Review 85*(9), 18–19.

Liu, L. and E. Yu (2004, April). Designing information systems in social context: A goal and scenario modelling approach. *Information Systems 29*(2), 187–203.

Mouton, F., M. M. Malan, K. K. Kimppa, and H. Venter (2015). Necessity for ethics in social engineering research. *Computers & Security 55*, 114–127.

Mussbacher, G., S. Ghanavati, and D. Amyot (2009). Modeling and Analysis of URN Goals and Scenarios with jUCMNav. In *Proceedings of the 2009 17th IEEE International Requirements Engineering Conference, RE*, RE '09, Washington, DC, USA, pp. 383–384. IEEE Computer Society.

Pendse, S. G. (2011). Ethical Hazards: A Motive, Means, and Opportunity Approach for Curbing Corporate Unethical Behavior. *Journal of Business Ethics 107*(3), 265–279.