PUBLIC PERCEPTIONS OF RE-IDENTIFICATION ATTACKS

Ester Moher; Khaled El Emam
Children's Hospital of Eastern Ontario; Privacy Analytics
401 Smyth Rd
Ottawa, ON

1. INTRODUCTION

Re-identification attacks, attacks where individuals within an anonymous data set are made identifiable by linking or other measures, are generally viewed as negative events. Reactions to a re-identification attack can include a financial settlement for individuals included in the dataset, financial and temporal costs of upgrading security systems to the organization, and other social costs such as brand defamation, reduced trust, and customer discontent. For this reason, some [1]–[4] have argued for more stringent legislation for adversaries attempting to re-identify data.

However, there has also been a call for greater sharing of anonymized data, for research and other purposes [5]. For example, HITRUST has developed a framework for sharing data secondarily, indicating implicit support for increased sharing of anonymized data [6]. Thus, there is a conflict between a desire to share data, and regulating how data is anonymized and used.

One factor in this debate relies on how the public perceives these re-identification attacks. Very little research has examined how judgments of harm from a re-identification attack are determined, and whether these judgments vary depending on contextual variables that shift from case to case. The purpose of the current research is to understand which types of re-identification attack are perceived to be riskier and more harmful, which may inform policy on data sharing.

2. METHODS

We examine whether certain elements of a re-identification attack influence how harmful the attack is perceived, in terms of patient trust, subjective harm to the patient, and need for punishment of attacker. We examine how individuals react when a hypothetical re-identification attack has been attempted on data that includes their own de-identified responses.

2.1 Variables of interest

The first variable of interest focuses on whether a re-identification attack may be considered more harmful when it is attempted by a black-hat organization relative to a white-hat organization.

> H1: Judged harmfulness of an attack will be greater when attackers are black-hat (nefariously motivated) versus when they are white-hat (virtuously motivated).

A second variable of interest examines whether the data that is attacked is managed privately or publicly. Previous research suggests the type of firm involved may influence both trustworthiness ratings and disclosure behaviour [7]–[10].

H2: Judged harmfulness of attack will be greater when the registry is private, relative to when it is public.

A third variable of interest is whether an internal review board (IRB) was involved in data management. IRB oversight can be seen as a means of providing assurances over data's safety and security, even in the event of an attack.

H3: Presence of an IRB will reduce judged harmfulness of the attack, especially when the firm is private versus public.

A fourth variable of interest focused on who the adversary was. Previous research [11]–[13] suggests that individuals feel more comfortable in sharing data with organizations perceived as "in-groups", or groups in which both the organization and the individual share an identity (relative to those perceived as "out-groups", or groups where no aspect of identity is shared).

H4: Judged harmfulness of breach will be greater when attackers are from out-groups relative to from in-groups.

The fifth and final variable of interest involved what type of information was attacked. Re-identification attacks on information that is more sensitive in nature may be seen as more harmful.

H5: Attacks will be considered more harmful when datasets contain individual-level, sensitive information, relative to when data sets are aggregate or contain non-sensitive information.

Participants were given a brief overview of re-identification attacks, and were then asked to read vignettes and complete a set of questions for each, assessing harm caused by each attack. We created a common template describing an attack, and varied each of 5 variables of interest, described below, for each possible pairing, resulting in 32 unique vignettes. Each participant received a randomly-allocated subset of six vignettes for brevity. As such, our design was a mixed within- and between-subjects design.

3. RESULTS

A sample of 106 participants were recruited from Crowdflower. Each participant completed 6 vignette scenarios.

Harm was judged to be greater when the attacker was part of a black hat organization (having nefarious motives), relative to when the attacker was part of a white hat organization (having virtuous motives; H1), $F(1, 621) = 50.48$, $MSE = 1.71$, $p < .01$, $\eta^2 = .08$. Further, harsher punishments were prescribed when attacker was a representative of a black hat organization, relative to a white hat organization, $F(1,621) = 62.53$, $p < .01$, $\eta^2 = .10$.

We did not observe any effects with regard to whether the attacked organizations were public institutions (such as academic research labs) or private firms (such as pharmaceutical companies; H2).

Data custodians were judged to be *more* responsible for an attack when an IRB oversaw data collection and storage, relative to when there was no ethics board oversight (counter to H3), $F(1,621) = 4.31$, $MSE = 3.30$, $p = .04$, $\eta^2 = .01$.

Participants thought the attacker should be punished more harshly when the attacker was from an out-group than from an in-group (H4), $F(1,621) = 2.78$, $MSE = 2.61$, $p = .10$, $\eta^2 = .01$.

Finally, sensitive, individual-level information was rated as more sensitive than was information that was aggregated or demographic in nature (H5), $F(1, 621) = 40.14$, $MSE = 3.01$, $p < .01$, $\eta^2 = .06$.

4. CONCLUSIONS

Findings of the current study suggest that the attacker's motivations were the primary determinant of how individuals react to an attack. As such, a white-hat attack, or an attack that allows for a quick patch to a problem, may be well worth publicizing. This publicity is likely to reduce patient concern, and may have fewer downstream consequences to the patient-provider relationship. However, organizations holding data should be aware that presence of an IRB or ethics committee may lead patients to heighten their standards. As such, organizations already in possession of such oversight must be especially cautious in disclosing re-identification attacks.

5. ACKNOWLEDGMENTS

We would like to thank the EHIL lab for insightful comments and suggestions, and to the SOUPS review committee for comments on a previous version of this work.

6. REFERENCES

[1] D. C. Barth-Jones, "Public Policy Considerations for Recent Re-Identification Demonstration Attacks on Genomic Data Sets: Part 1 (Re-Identification Symposium)," 29-May-2013. .

[2] D. C. Barth-Jones, "Press and Reporting Considerations for Recent Re-Identification Demonstration Attacks: Part 2 (Re-Identification Symposium)," 01-Oct-2013. .

[3] D. C. Barth-Jones, "Ethical Concerns, Conduct and Public Policy for Re-Identification and De-identification Practice: Part 3 (Re-Identification Symposium)," 02-Oct-2013. .

[4] S. R. Gelman, D. Pollack, and A. Weiner, "Confidentiality of Social Work Records in the Computer Age," *Social Work*, vol. 44, no. 3, pp. 243–252, 1999.

[5] Federal Trade Commission, "Protecting consumer privacy in an era of rapid change," FTC Report, Mar. 2012.

[6] HITRUST, "HITRUST to Improve Patient Privacy with New Framework for De-Identification of Health Information," 12-Mar-2015. [Online]. Available: https://hitrustalliance.net/hitrust-improve-patient-privacy-new-framework...

[7] B. P. Knijnenburg, "Information Disclosure Profiles for Segmentation and Recommendation," presented at the Symposium on Usable Privacy and Security (SOUPS), Menlo Park, CA., 2014, p. 4.

[8] A. N. Joinson and C. B. Paine, "Self-disclosure, Privacy and the Internet," in *The Oxford handbook of Internet psychology*, 2007, pp. 237–252.

[9] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. Paine Schofield, "Privacy, Trust And Self-Disclosure Online," *Human-Computer Interaction*, vol. 25, no. 1, pp. 1–24, 2010.

[10] A. E. G. Skinner and G. Latchford, “Attitudes to counselling via the Internet: A comparison between in-person counselling clients and Internet support group users,” *Counselling and Psychotherapy Research*, vol. 6, no. 3, pp. 158 –163, 2006.

[11] M. B. Brewer, “In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis,” *Psychological Bulletin*, vol. 8, no. 2, pp. 307–324, 1979.

[12] M. B. Brewer and W. Gardner, “Who Is This ‘We’? Levels of Collective Identity and Self Representations,” *Journal of Personality and Social Psychology*, vol. 71, no. 1, pp. 83–93, 1996.

[13] H. Tajfel, M. G. Billig, R. P. Bundy, and C. Flament, “Social categorization and intergroup behaviour,” *Eur. J. Soc. Psychol.*, vol. 1, no. 2, pp. 149–178, 1971.