Analyzing 4 Million Real-World Personal Knowledge Questions (Short Paper)

Maximilian Golla and Markus Dürmuth

Horst Görtz Institute for IT-Security, Ruhr-University Bochum, Germany
{maximilian.golla,markus.duermuth}@rub.de

Abstract. Personal Knowledge Questions are widely used for fallback authentication, i. e., recovering access to an account when the primary authenticator is lost. It is well known that the answers only have low-entropy and are sometimes derivable from public data sources, but ease-of-use and supposedly good memorability seem to outweigh this drawback for some applications.
Recently, a database dump of an online dating website was leaked, including 3.9 million plain text answers to personal knowledge questions, making it the largest publicly available list. We analyzed this list of answers and were able to confirm previous findings that were obtained from non-public lists (WWW 2015), in particular, we found that some users don't answer truthfully, which may actually reduce the answer's entropy.

Keywords: fallback authentication, personal knowledge question, password recovery, password reset, challenge question

1 Introduction

Personal Knowledge Questions (PKQ) are commonly used for fallback authentication, i. e., to recover access to an account when the primary authenticator is lost. Common examples of personal knowledge questions ask for easy-to-remember facts about a user's life, such as the "mother's maiden name" or the "brand of the first car". Previous studies indicated that these questions have indeed better memorability [17], where likely explanations are that (i) no secret needs to be remembered, and (ii) it bases on a cued-recall problem (where the question asked constitutes the cue), in contrast to, e. g., traditional passwords that base on a (pure) recall problem. However, a number of problems limit the usefulness of PKQs: (i) In some instances, the answers to PKQs are available in public databases [7, 13], (ii) answers may be easily guessable by friends and relatives [15], and (iii) the entropy of answers may be low, due to frequent answers or small answer space [15, 4, 13]. In the past, the study of PKQs was hindered by a lack of publicly available real-world data, the only exception being recent work by Bonneau et al. [3], which studied personal knowledge questions at Google, with the data not being publicly available.

In August 2015, a database dump of an online dating service called Ashley Madison was leaked [11]. This data includes about 3.9 million answers to PKQs, which makes it by far the largest set publicly available. The leak is widely considered authentic and contains answers to four different questions (Mother's Maiden Name, Name of High School, Name of Favorite Sports Team, Last 4 Digits of SSN), with 300,000 to 1,500,000 answers per question. In this work, we provide a first analysis of this data, where we focus on the guessability of the answers. We use statistical entropy measures (partial guessing entropy [3]), which assumes perfect knowledge of the distribution of answers, and thus provides a lower-bound on the security offered. Our findings include:

i) The security of knowledge questions is low compared to other knowledge-based authentication methods.
ii) The country of origin and the age of the user influences the strength of the answer for some questions but not for others.
iii) The question about the sports team is the least secure one.

1.1 Related Work

One of the earlier studies on the security of PKQs was conducted in 1993 [17], which found good usability and security. However, by today's standards, the security of PKQs is rather low, as several studies have shown. Griffith et al. showed [7] that the *Mother's Maiden Name*, which is frequently used as a PKQ, can often be derived from public databases, rendering them insecure. Rosenblum has shown that private information about persons can often be inferred from social networking sites [14]. This information can be used to narrow down potential answers for the security questions. Secrecy of those answers in the age of Facebook was studied by Rabkin [13], whereas Bonneau et al. studied the entropy of names [4]. Schechter et al. demonstrated [15] that for a number of such questions the answers can often be guessed easily. A more general discussion on designing security questions including usability, privacy, and security is given by Just [9]. Alternative and a potentially better domain of security questions, namely questions about *personal preference* similar to those used on online dating sites, where studied by Jakobson et al. [8] and have found to provide better security than most other commonly used questions. In recent work, Bonneau et al. [3] have evaluated real-world PKQs at Google and found that some answers are quite predictable, in part because some users do not answer truthfully, which indeed lowers the overall security. This work is the closest to our work; the biggest difference being that our work is based on a public data source and thus is reproducible.

Alternative forms of fallback authentication use a registered email address or mobile phone of the user [6], where an access code is sent to the registered device if the user lost the regular password and requested a password reset. Fallback authentication by support teams is susceptible to social engineering attacks [12]. Social authentication, or vouching, was proposed by [5], a more recent design is given by Schechter et al. [16]. Problems with social authentication were pointed out in [10].

2 Methodology

Attacker Model and Guessing Metrics: We consider an attacker guessing the answer of PKQs without specific knowledge about the user. We consider an idealized attacker that has exact knowledge about the distribution of answers to the questions, modeled by statistical guessing metrics [2]. We are mostly interested in *online guessing attacks*, where we assume that only a very limited number of guesses can be made before the account is locked, or substantial delays are applied to slow down guessing attempts. Resistance against online guessing attacks can be measured by the fraction λ_β of accounts that covers the β most common answers, i. e., an attacker that can make β guesses before being locked out can compromise roughly λ_β of the accounts. For comparability, we follow Bonneau et al. [3] and also list partial guessing entropy G_α, which measures the resistance against offline guessing attacks.

Datasets: The dating website used four different security questions for password recovery: *What is Your Mother's Maiden Name?* (MMN), *What is the Name of Your High School?* (High School), *What is Your Favorite Sports Team?* (Sport Team), and *What are the Last 4 Digits of Your SSN?* (SSN). In total, the dataset contained approx. 32 million entries, where 3.9 million had a security answer set.

It has been noted earlier by Newitz [1] that the leaked dataset contained entries from bots, which were used to chat with users that had few "real" contacts. She found a set of criteria to approximately differentiate between datasets that belong to humans and those that belong to bots. Those criteria concern the used email addresses, indicators for the last contact by mail or chat, IP addresses, as well as the existence of account deletion flags within the password hash strings. After filtering for bots, there were 903,255 entries remaining for MMN (out of 1,576,779 before filtering), 632,484 entries for High School (out of 1,031,416), 650,680 entries for Sports Team (out of 1,011,383), and 186,134 entries for SSN (out of 309,827).

Ethical Considerations: The data analyzed in this paper was leaked to the public before, so attackers already had independent access to the datasets. We took care that the data was not distributed any further, and we only present aggregated results in this work that will not leak information about any specific answer.

3 Strength Evaluation

Table 1 shows the full results for all subsets we consider. It lists both λ_β as well as guessing entropy values G_α, with the same parameters used by Bonneau et al. [3] so that the results are comparable to their work.

Sample Size and Significance: To test for statistical significance of our results, we use a slightly simplified re-sampling approach similar to Bonneau et al. [3]. For each set, we repeatedly sample a random subset of 10% the original size and

			online guessing (success in %)					offline guess. (bits)		
		size	λ_1	λ_3	λ_{10}	λ_{100}	λ_{1000}	$\tilde{G}_{0.1}$	$\tilde{G}_{0.25}$	$\tilde{G}_{0.5}$
mother's maiden name										
all		903 255	3.21	5.12	8.18	22.41	48.10	7.34	8.93	10.94
country	USA	612 890	*3.15*	5.31	8.89	24.41	51.74	7.07	8.64	10.42
	Canada	125 101	-	*4.09*	*6.91*	*21.64*	*48.49*	7.68	9.03	10.86
	UK	26 912	-	-	-	-	-	*6.27*	*7.97*	*9.53*
age	<= 35	169 571	*2.65*	*4.06*	*6.79*	20.73	*46.37*	7.79	9.19	11.18
	36 − 45	293 868	*3.08*	*5.12*	8.08	22.63	48.33	7.35	8.90	10.90
	46 − 55	284 594	*3.48*	*5.55*	8.82	23.38	49.36	7.08	8.79	10.75
	> 55	155 222	*3.58*	*5.50*	*8.91*	23.74	*49.78*	7.07	8.74	10.69
high school										
all		632 484	2.63	5.60	7.86	16.91	37.88	7.78	10.04	11.93
country	USA	461 582	*2.68*	6.03	8.55	18.25	42.61	7.36	9.67	11.39
	Canada	86 465	-	-	*9.83*	*26.79*	*63.56*	6.72	8.32	9.59
	UK	8 740	-	-	-	-	-	-	-	-
age	<= 35	127 707	*2.56*	*5.76*	*7.92*	*16.45*	-	7.79	10.30	*12.28*
	36 − 45	217 157	*2.57*	5.51	7.72	16.88	*37.90*	7.85	10.04	11.92
	46 − 55	194 608	*2.80*	*5.53*	7.87	17.27	*40.21*	7.72	9.90	11.63
	> 55	93 012	-	*5.86*	*8.39*	*18.60*	-	7.43	9.63	11.25
sports team										
all		650 680	2.83	7.84	19.44	62.94	89.68	5.39	5.82	6.58
country	USA	432 129	4.11	10.75	26.51	74.15	93.08	4.79	5.20	5.75
	Canada	85 520	*11.16*	18.61	36.03	74.61	91.44	3.16	4.39	5.30
	UK	12 011	-	*23.23*	*41.93*	*73.49*	-	*3.50*	*3.87*	*4.86*
age	<= 35	128 520	-	*6.03*	16.38	58.78	86.98	5.75	6.16	6.90
	36 − 45	250 901	*2.73*	7.69	19.73	63.39	89.91	5.43	5.77	6.53
	46 − 55	199 321	*3.07*	8.67	21.94	65.83	91.40	5.15	5.60	6.35
	> 55	71 938	-	*9.48*	23.49	66.95	*91.73*	5.05	5.48	6.25
SSN (4 digits)										
all		186 134	-	*5.32*	*7.07*	-	-	*9.91*	*12.15*	*12.70*
country	USA	128 611	-	*5.15*	*6.73*	-	-	-	*12.17*	*12.65*
baseline										
password (RockYou) [3]			0.9	1.4	2.1	4.6	11.3	12.8	15.9	19.8
4-digit PIN (iPhone) [3]			4.3	9.2	14.4	29.3	56.4	5.2	7.7	10.1
4-digit random PIN			0.01	0.03	0.1	1.0	10.0	13.29	13.29	13.29

Table 1. Full listing of the results.

test if the resulting entropy values fall within a 5% or a 10% error band with a confidence of at least $p = 0.98$. Values that are in the 5% band are shown in non-italic, those that fall in the 10% band are shown in italic. All other values are omitted from the table.

Differences in Questions: First, we compare the four different questions available. We found that the results for λ_1 are surprisingly consistent over all tested questions and subsets. However, higher values, e. g., λ_{10}, show measurable differences. For example, we have $\lambda_{10} = 19.44$ for the sports team question, while the other three questions all have a λ_{10} of around 8. This is also reflected in the guessing entropy, with a $\tilde{G}_{0.25}$ of 5.82 for sports team, and 12.15 for the SSN.

Differences in Nationality: When considering different nationalities, we observe that mother's maiden name and high school show surprisingly little variation. For sports team, interestingly, knowing the nationality has a substantial influence on the guessability of the answer (we have λ_1 of 2.83 if nationality is unknown, and values between 04.11 and 11.16 once the nationality is known). In addition, for Canada, the sports teams are slightly easier to guess than for the US. The SSN is substantially less secure when a person is from outside the US. This is most likely explained by the fact that the SSN is specific to the US, and answers from non-US citizens contain a substantial fraction of fake answers, a behavior which was similarly observed in previous work [3]. For the complete set of the SSN each answer should occur, assuming a uniform distribution, with probability 0.01%, but the most popular answer (1234) occurred with probability 2.23%. Other frequent answers include 1111, 0000, and 6969.

Differences in Age: The data suggests that the age has little influence on the guessability; for mother's maiden name and SSN the differences are practically non-existent. For high school and sports team, however, we find that with increasing age the answers are slightly easier to guess, which means that there is less variability (e. g., for sports team we have $\lambda_{10} = 16.38$ for those under 35 and $\lambda_{10} = 23.49$ for those over 55).

Baseline: As a baseline, we added entropy values for passwords (using the RockYou list), for real-world PINs, and for randomly chosen PINs; the entropy values for the first and second one are taken from [3]. These provide an interesting perspective: We see that all knowledge questions in the dataset are substantially less secure than a randomly chosen 4-digit PIN, however, they are slightly more secure than a real-world PIN chosen by a human.

4 Conclusion

We have analyzed the answers to personal knowledge questions given in the data leaked from Ashley Madison in August 2015. We found that favorite sports teams are particularly easy to guess, that the security depends, to a certain extent, on the age and the origin of a user, and that in general they only offer a low level of security.

References

1. Annalee Newitz – Gawker Media. Ashley Madison Code Shows More Women, and More Bots, August 2015. Online at http://gizmodo.com/ashley-madison-code-shows-more-women-and-more-bots-1727613924, as of August 3, 2016.
2. Joseph Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*. IEEE, 2012.
3. Joseph Bonneau, Elie Bursztein, Ilan Caron, Rob Jackson, and Mike Williamson. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In *International World Wide Web Conference*. IW3C2, 2015.
4. Joseph Bonneau, Mike Just, and Greg Matthews. What's in a Name? Evaluating Statistical Attacks on Personal Knowledge Questions. In *Financial Cryptography and Data Security*, volume 6052 of *LNCS*, pages 98–113. Springer, 2010.
5. John Brainard, Ari Juels, Ronald L. Rivest, Michael Szydlo, and Moti Yung. Fourth Factor Authentication: Somebody You Know. In *ACM Conference on Computer and Communications Security*, pages 168–178. ACM Press, 2006.
6. Simson L. Garfinkel. Email-Based Identification and Authentication: An Alternative to PKI? *IEEE Security and Privacy*, 1(6):20–26, 2003.
7. Virgil Griffith and Markus Jakobsson. Messin with Texas: Deriving Mothers Maiden Names Using Public Records. In *Applied Cryptography and Network Security*, volume 3531 of *LNCS*, pages 91–103. Springer, 2005.
8. Markus Jakobsson, Erik Stolterman, Susanne Wetzel, and Liu Yang. Love and Authentication. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200. ACM Press, 2008.
9. Mike Just. Designing and Evaluating Challenge-Question Systems. *IEEE Security and Privacy*, 2(5):32–39, 2004.
10. Hyoungshick Kim, John Tang, and Ross Anderson. Social Authentication: Harder Than It Looks. In *Financial Cryptography and Data Security*, volume 7397 of *LNCS*, pages 1–15. Springer, 2012.
11. Kim Zetter – Wired. Hackers Finally Post Stolen Ashley Madison Data, August 2015. Online at http://www.wired.com/2015/08/happened-hackers-posted-stolen-ashley-madison-data/, as of August 3, 2016.
12. Kevin D. Mitnick and William L. Simon. *The Art of Deception: Controlling the Human Element of Security.* John Wiley & Sons, 2002.
13. Ariel Rabkin. Personal Knowledge Questions for Fallback Authentication: Security Questions in the Era of Facebook. In *USENIX Symposium on Usable Privacy and Security*, pages 13–23. USENIX Association, 2008.
14. D. Rosenblum. What Anyone Can Know: The Privacy Risks of Social Networking Sites. *IEEE Security and Privacy*, 5(3):40–49, 2007.
15. Stuart Schechter, A. J. Bernheim Brush, and Serge Egelman. It's No Secret. Measuring the Security and Reliability of Authentication via "Secret" Questions. In *IEEE Symposium on Security and Privacy*, pages 375–390. IEEE Computer Society, 2009.
16. Stuart Schechter, Serge Egelman, and Robert W. Reeder. It's Not What You Know, But Who You Know: A Social Approach to Last-Resort Authentication. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 1983–1992. ACM Press, 2009.
17. M. Zviran and W. J. Haga. A Comparison of Password Techniques for Multilevel Authentication Mechanisms. *The Computer Journal*, 36(3):227–237, 1993.